

Sharing knowledge: scientific communication: 9. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaft in Deutschland

Veröffentlichungsversion / Published Version
Konferenzband / conference proceedings

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Informationszentrum Sozialwissenschaften; IuK - Initiative Information und Kommunikation; Deutsche Mathematiker-Vereinigung e.V. (DMV); Institut für Wissenschaftliche Information (IWI). (2004). *Sharing knowledge: scientific communication: 9. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaft in Deutschland* (Tagungsberichte / GESIS-IZ Sozialwissenschaften, 8). Bonn. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-263421>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.



Sharing Knowledge: Scientific Communication

**9. Kongress der luK-Initiative der
Wissenschaftlichen Fachgesellschaft in Deutschland**

Tagungsberichte Band 8



InformationsZentrum
Sozialwissenschaften

GESIS

Sharing Knowledge: Scientific Communication

Tagungsberichte

Herausgegeben vom Informationszentrum Sozialwissenschaften (IZ)
der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI), Bonn.
Band 8

Das IZ ist Mitglied der Gesellschaft Sozialwissenschaftlicher
Infrastruktureinrichtungen e.V. (GESIS).
Die GESIS ist Mitglied der Leibniz-Gemeinschaft.

IuK - Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland:
<http://www.iuk-initiative.org/>

IWI Institut für Wissenschaftliche Information e.V. (Osnabrück):
<http://www.iwi-iuk.org/>

Ausrichtung der Tagung unterstützt von: Deutsche Mathematiker Vereinigung:
<http://www.mathematik.uni-bielefeld.de/dmv/mathnet.html>

Sharing Knowledge: Scientific Communication

9. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaft in Deutschland

IuK-Initiative Information und Kommunikation
Deutsche Mathematiker Vereinigung (DMV), Berlin
Institut für Wissenschaftliche Information (IWI), Osnabrück

Tagungsbericht Band 8

Informationszentrum Sozialwissenschaften, Bonn 2004

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Sharing Knowledge: Scientific Communication -
Bonn : IZ Sozialwiss., 2004 (Tagungsberichte ; Bd. 8)
ISBN 3-8206-0145-7

Herausgeber,

Druck und Vertrieb: Informationszentrum Sozialwissenschaften
Lennéstraße 30, 53113 Bonn
Tel.: 02 28 - 22 81 - 0
Printed in Germany

© 2004 Informationszentrum Sozialwissenschaften, Bonn. Alle Rechte vorbehalten. Insbesondere ist die Überführung in maschinenlesbare Form sowie das Speichern in Informationssystemen, auch auszugsweise, nur mit schriftlicher Einwilligung gestattet.

Inhalt

Vorwort

.....	9
-------	---

Entwicklung von alternativen Publikationsstrukturen in Europa und den USA

DFG-Projekt: Perspektiven für den Bezug elektronischer Fachinformation in der Bundesrepublik Deutschland

<i>Heike Andermann</i>	11
------------------------------	----

Qualitätssicherung und Nutzung von wissenschaftsrelevanten Angeboten auf dem Deutschen Bildungsserver: Empirische Befunde einer vergleichenden Logfile-Analyse

<i>Elke Brenstein, Olaf Kos</i>	21
---------------------------------------	----

vascoda

Das gemeinsame Portal von Informationsverbünden, Elektronischer Zeitschriftenbibliothek und Virtuellen Fachbibliotheken

<i>Christine Burbliès, Tamara Pianos</i>	37
--	----

Probleme der Integration digitaler Bibliothekssysteme: Semantische Heterogenität bei datenbankübergreifenden Recherchen

<i>Klaus Hahn</i>	47
-------------------------	----

Unterstützung kooperativer Verfahren zum Aufbau von Fachportalen

<i>Heiko Hellweg, Bernd Hermes und Maximilian Stempfhuber</i>	59
---	----

PhysNet und seine Spiegel - Das Projekt SINN

<i>Michael Hohlfeld, Thomas Severiens</i>	71
---	----

Online-Hochschulschriften für die Praxis

<i>Nikola Korb</i>	83
--------------------------	----

Infrastrukturen für innovative Digital Library-Dienste Perspektiven des Kompetenznetzwerks „Neue Dienste, Standardisierung, Metadaten“ <i>Rudi Schmiede, Stephan Körnig</i>	95
Das Psychologische Datenarchiv PsychData Zur Frage der angemessenen Dokumentation von Primärdaten der empirisch-psychologischen Forschung <i>Jutta von Maurice</i>	107
Maßnahmen zur Förderung der Informationskompetenz durch den Fachinformationsanbieter ZPID <i>Astrid Nechvátal</i>	119
LIMES – A System for a Distributed Database Production in Mathematics <i>Olaf Ninnemann, Bernd Wegner</i>	131
Integration von digitalen Bibliotheken in Learning-Management-Systeme <i>Frank Oldenettel, Michael Malachinski</i>	139
MPRESS - transition of metadata formats <i>Judith Plümer</i>	157
Einsatz von Lernmanagementsystemen in traditionellen Lehrveranstaltungen am Beispiel der Universitäten Oldenburg und Osnabrück Ein Erfahrungsbericht <i>Dennis Reil</i>	169
Reflections on the Value Chain of Scientific Information <i>Hans E. Roosendaal, Peter A. Th. M. Geurts</i>	181
SozioNet -Web-Ressourcen für ein künftiges Fachportal Sozialwissenschaften <i>Natascha Schumann, Wolfgang Meier, Sue Heise, Rudi Schmiede</i>	193

ViFaPhys – Virtuelle Fachbibliothek Physik Ein kooperatives Dienstleistungsangebot für die Physik <i>Thomas Severiens, Esther Tobschall</i>	205
Weiterentwicklung von digitalen Bibliothekssystemen zu OpenArchives-Systemen <i>Günther Specht, Michael G. Bauer</i>	213
Zeitliche Verwaltung XML-basierter Metadaten in digitalen Bibliotheken <i>Markus Kalb, Günther Specht</i>	227
infoconnex Der Informationsverbund Pädagogik – Sozialwissenschaften – Psychologie <i>Maximilian Stempfhuber</i>	239
Das didaktische Metadaten-System DML Grundlagen und praktische Anwendung <i>Christian Swertz</i>	249
The C₂M project: a wrapper generator for chemistry and biology <i>Paul van der Vet, Eelco Mossel</i>	267
Analyse der Qualität der multimedialen Lernumgebung „BioPrinz“ <i>Josef Wiemeyer</i>	283

Vorwort

Zur neunten Frühjahrstagung der Initiative Information und Kommunikation wissenschaftlicher Fachgesellschaften (IuK) erscheint erstmals zusätzlich zu den Webproceedings ein referierter Tagungsband.

Aus gut der Hälfte der Vorträge der Tagung sind letztlich Beiträge zum Tagungsband hervorgegangen.

Die Webproceedings, die die Dokumentation der Powerpoint-Folien der Vorträge enthalten, sind unter <http://www.iwi-iuk.org/iuk2003/> zugänglich.

Mit dem Tagungsband unternimmt die Initiative einen konkreten Schritt zur Verfolgung von Zielsetzungen ihres Grundsatzpapiers zu digitalen Bibliotheken. Sie intensiviert den Diskurs über die durch neuere Techniken gestützte Kommunikation innerhalb und zwischen wissenschaftlichen Gemeinschaften. Die Form des referierten Artikels soll dabei die (selbst)-kritische Auseinandersetzung mit dem in Projekten Angestrebten und Erreichten und die Überprüfung „visionärer“ Entwürfe am Realen fördern.

In diesem wissenschaftlichen Diskurs finden Platz: Entwicklung bei den Techniken als solchen, Entfaltung neuer Materialtypen, nachhaltige Sicherung des Betriebs von Installationen („Geschäftsmodelle“), fachliche Spezifität und Expressivität sowie Skalierbarkeit im inter- und überdisziplinären Rahmen.

Der Tagungsband stellt auch eine Momentaufnahme des Entwicklungsstandes digitaler Bibliotheken in der Bundesrepublik dar und der in dieser Entwicklung noch nicht beantworteten und noch nicht gestellten Fragen.

Mit Dank an die Autoren, das Programmkomitee, die Referenten sowie die geduligen und verständnisvollen Mitarbeiter des IZ Sozialwissenschaften Bonn.

Roland Schwänzl
IuK und IWI e.V.

Entwicklung von alternativen Publikationsstrukturen in Europa und den USA

DFG-Projekt: Perspektiven für den Bezug elektronischer Fachinformation in der Bundesrepublik Deutschland¹

Heike Andermann

Universitätsbibliothek Potsdam

1 Krise der wissenschaftlichen Informationsversorgung

Im Übergang vom analogen zum digitalen Medium verschärft sich der Konflikt zwischen Urhebern (WissenschaftlerInnen), Rechteverwertern (Verlagen) und institutionellen Abnehmern (Bibliotheken) der wissenschaftlichen Information. Es fehlen medienadäquate Geschäftsmodelle für den Umgang mit digitalen Dokumenten, rechtliche Rahmenbedingungen für die Verbreitung und Nutzung wissenschaftlicher Ergebnisse in elektronischen Umgebungen und wissenschaftsfreundliche Preisgestaltungen für Fachzeitschriften. Für die Preisentwicklungen bei den Fachzeitschriften hat sich mittlerweile der Begriff der Krise der wissenschaftlichen Informationsversorgung durchgesetzt. Gemeint ist eine steigende Lücke zwischen dem Preis für die wissenschaftliche Information und der Zahlungsfähigkeit der Bibliotheken mit der Konsequenz, das nicht nur Zeitschriften, sondern auch Monographien, insbesondere in den nicht-naturwissenschaftlichen Fachgebieten abbestellt werden. Ursächlich verantwortlich für diese Entwicklung sind mehrere Faktoren, die hier nur kurz benannt werden sollen: eine zunehmende Kommerzialisierung der STM²-Information, Konzentrationsprozesse im Informationsmarkt, die die Durchsetzung von Preiserhöhungen für Fachzeitschriften vereinfachen³, restriktive Geschäftsmodelle im Übergang zum digitalen Medium sowie eine generelle Zunahme des Angebotes an Fachzeitschriften, welches eine höhere Nachfrage nach sich zieht. Vom Verband wissenschaftlicher Bibliotheken in Nordamerika (ARL) wird die Situation folgendermaßen beschrieben: „Expenditures for electronic serials have increased by 75% in the last two years alone, and by almost 900% since they were first repor-

1 Weitere Informationen zu dem Projekt finden Sie unter: <http://www.epublications.de>

2 Science, Technology and Medicine

3 vgl. hierzu ausführlicher McCabe, Mark (1999): The Impact on Publisher Mergers on Journal Prices. A Preliminary Report. <http://www.arl.org/newsltr/200/mccabe.html>

ted, in 1994-1995.⁴ Beispielhaft für die europäische Situation ist die Einschätzung der Preisentwicklungen in den englischen Bibliotheken. Von der Universität Liverpool werden die Entwicklungen folgendermaßen beschrieben: „Journals purchased by UK academic libraries have gone up in price an average of 11% per annum since 1991-92, while the consumer price index has increased by only 2,7% per annum. Academic libraries in the UK spent 19% more per FTE student to purchase 18% fewer journal titles per FTE student in 1999-2000 than in 1991-92.“⁵

Mit der zunehmenden Akzeptanz der digitalen Medien erscheinen die für die „Printwelt“ entwickelten rechtlichen Rahmenbedingungen immer weniger zeitgemäß. Diese sichern den Verlagen das exklusive Recht für die Verwertung und Verbreitung der Zeitschriftenprodukte⁶ zu. Die Verlagspolitik besteht darin, nur solche Artikel zu veröffentlichen, die in einer anderen Zeitschrift noch nicht erschienen sind. Verbreitet sind sogenannte „buy-out-Verträge“, bei dem der Urheber sämtliche Nutzungsarten (sachlich, zeitlich, räumlich) gegen eine einmalige Zahlung an den Werkverwerter überträgt⁷ mit der Konsequenz, dass die Artikel in keiner anderen Zeitschrift bzw. in keinem anderen Medium veröffentlicht werden dürfen.

Unter den derzeitigen Bedingungen werden die Innovationspotenziale, die sich mit dem Einsatz der neuen Produktions- und Verbreitungstechnologien für die Neugestaltung des wissenschaftlichen Kommunikationsprozesses ergeben, nicht annähernd ausgeschöpft. Digitalisierung ermöglicht die Vervielfältigung des Dokumentes, ohne die Qualität zu beeinträchtigen. Die Internettechnologien ermöglichen eine rasche und globale Verbreitung der wissenschaftlichen Inhalte.⁸ Nur durch eine Veränderung der bestehenden Geschäftsmodelle und der rechtlichen Rahmenbedingungen gelingt es in Zukunft, die Möglichkeiten der neuen Technologien für den wissenschaftlichen Kommunikationsprozess nutzbar zu machen. Gegenwärtig sind auf dem Feld des elektronischen Publizierens Initiativen und Unternehmungen zu beobachten, die darauf abzielen, den Prozess wissenschaftlicher Kommunikation auf der Basis dieser neuen

4 <http://www.arl.org/stats/pubpd/sup01.pdf>

5 <http://www.liv.ac.uk/Library/createchange/liverpoolcontext.html>

6 Diese Vorgehensweise wird auch als „Ingelfinger-Gesetz“ nach dem Herausgeber der Zeitschrift *New England Journal of Medicine* Franz Ingelfinger benannt. Vgl. ausführlicher: Harnad, Stevan (2000): *Ingelfinger Over-Ruled: The Role of the Web in the Future of Refereed Medical Journal Publishing*. Preprint. <http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad00.lancet.htm>

7 vgl. hierzu ausführlicher Schröter, Madeleine (2002): Der (Copyright-)Vertrag des Wissenschaftlers mit dem Verlag. In: *die Zukunft des wissenschaftlichen Publizierens: Der Wissenschaftler im Dialog mit Verlag und Bibliothek*, Schriftenreihe des Forschungszentrums Jülich, Bd.10, S.42ff.

8 vgl. Kühlen, Rainer (2002): *Medienprodukte im Netz – Zwischen Kommerzialisierung und freiem Zugang*.

Technologien neu zu gestalten. Sie lassen sich in vier Aktionsfelder differenzieren:

- Freier Zugang zu wissenschaftlicher Information auf der Basis neuer Geschäftsmodelle
- Selbstorganisation der Wissenschaften - Aufbau verlagsunabhängiger Publikationsinfrastrukturen
- Neue Kooperationsformen für die Publikation und Verbreitung wissenschaftlicher Information
- Kostensenkung durch direkten Wettbewerb

2 Freier Zugang zu wissenschaftlicher Information auf der Basis neuer Geschäftsmodelle⁹

Kennzeichnend für das neue Geschäftsmodell im digitalen Kontext ist der *freie*, d.h. kostenlose Zugang zur wissenschaftlichen Information nach den Prinzipien der Budapest Open Access Initiative.¹⁰ Mit der Möglichkeit des globalen Zugriffs auf die wissenschaftliche Information wird eine zentrale Zielsetzung wissenschaftlichen Publizierens realisiert: die Maximierung der Sichtbarkeit und Aufmerksamkeit für die Forschungsergebnisse.

Die Finanzierung der Produktion und Verbreitung der frei zugänglichen Zeitschriften erfolgt nicht länger durch die Abnehmer der wissenschaftlichen Information, sondern durch die Urheber bzw. durch die Institution, der die Urheber angehören. Zur Finanzierung der verlegerischen Tätigkeiten werden Artikelbearbeitungsgebühren bzw. Autorenggebühren erhoben, deren Höhe derzeit zwischen US \$500¹¹ und US \$1500¹² variiert. Von der *Public Library of Science*¹³, Herausgeber zweier Fachzeitschriften in der Biologie und Medizin und *BioMed Central*¹⁴, einem Online-Verlag für die Fachgebiete Biologie und Medizin, wird dieses Finanzierungskonzept ergänzt durch die Möglichkeit einer institutionellen Mitgliedschaft von Hochschulen. Universitäten können für „ihre“ Wissen-

9 Weitere Beispiele: New Journal of Physics, Molecular Diversity Preservation International

10 Open access bedeutet, „... dass diese Literatur kostenfrei und öffentlich im Internet zugänglich sein sollte, so dass Interessierte die Volltexte lesen, herunterladen, kopieren, verteilen, drucken, in ihnen suchen, auf sie verweisen und sie auch sonst auf jede denkbare legale Weise benutzen können, ohne finanzielle, gesetzliche oder technische Barrieren jenseits von denen, die mit dem Internet-Zugang selbst verbunden sind.“ <http://www.soros.org/openaccess/g/index.shtml> (Zugriff: 04.04.03)

11 BioMedCentral, New Journal of Physics, Molecular Diversity Preservation International Foundation

12 PloS Biology, PloS Medicine

13 <http://www.publiclibraryofscience.org/>

14 <http://www.biomedcentral.com/>

schaftlerInnen eine Publikationsmöglichkeit in den verlagseigenen Zeitschriften erwerben. Bei *BioMed Central* machen derzeit etwa 100 Institutionen von diesem Angebot Gebrauch. Als zusätzlicher Finanzierungsbaustein wird Firmen Werbefläche auf den Seiten des Online-Verlages angeboten. Die Autoren behalten die Rechte an ihrer Publikation und bei *BioMedCentral* ist eine Veröffentlichung als Preprint möglich. Zahlreiche Zeitschriften dieses Verlages werden in diesem Jahr erstmals in die Zitationsindices des Institute of Scientific Information (ISI) aufgenommen.

3 Selbstorganisation der Wissenschaften – Aufbau verlagsunabhängiger Publikationsinfrastrukturen

Von der *Scholarly Publishing & Academic Resources Coalition* (SPARC) wurde im vergangenen Jahr ein Positionspapier herausgegeben, in dem der Aufbau wissenschaftseigener Publikationsinfrastrukturen empfohlen wird.¹⁵ Zahlreiche Hochschulen in Europa und den USA sind im Begriff, diese sog. institutionellen Repositorien in technischer und organisatorischer Hinsicht zu realisieren. Was zeichnet die elektronischen Archive aus? Sie basieren auf dem Ansatz der Selbstarchivierung und können Preprints, Working Papers, publizierte Artikel, Lehr- und Lernmaterialien, Qualifikationsarbeiten und Simulationen etc. beinhalten. Die Qualitätssicherung erfolgt durch die beteiligten WissenschaftlerInnen bzw. durch die akademischen Strukturen. Sie tragen zu einer Beschleunigung des wissenschaftlichen Kommunikationsprozesses bei und reduzieren die Kosten für die Verbreitung und den Zugang zur wissenschaftlichen Information. Mit dem Aufbau dieser Archive ist ein dauerhafter Zugriff auf die wissenschaftlichen Ergebnisse in Zukunft sichergestellt. Von den beteiligten Akteuren werden die institutionellen Repositorien als ein Indikator für die Qualität und Relevanz der wissenschaftlichen community und der Hochschule im internationalen Wettbewerb angesehen.

In den USA und Europa sind relativ weit vorangeschrittene Entwicklungen auf diesem Gebiet zu beobachten. Im vergangenen Jahr ist an der Universität von Kalifornien *eScholarship*¹⁶ „eröffnet“ worden. Dieses elektronische Archiv bietet Zugriff auf sämtliche Tools für das Management, die Veröffentlichung und Verbreitung von digitalen Dokumenten. In Kooperation mit einem Online-Verlag werden Publikationsmöglichkeiten in digitalen Zeitschriften mit einem Online-Begutachtungsprozess angeboten.

¹⁵ Vgl. hierzu auch die Empfehlungen der Hochschulrektorenkonferenz (2002): Zur Neuausrichtung des Informations- und Publikationssystems der deutschen Hochschulen. URL: http://www.hrk.de/downloads/Empfehlung_Bibliothek.pdf (Zugriff: 03.04.03)

¹⁶ <http://escholarship.cdlib.org/>

Ein weiteres US-amerikanisches Projekt ist *DSpace*¹⁷, eine Kooperation zwischen den Bibliotheken des Massachusetts Institute of Technology (MIT) und Hewlett Packard (HP), die eine open source Software für das Management dieses elektronischen Archivs entwickelt haben. Sämtliche Dokumentformate, Simulationen und multimediale Veröffentlichungen sind archivierbar. Die Bibliotheken bieten Dienstleistungen im Zusammenhang mit diesen elektronischen Dokumenten an. Hierzu zählen die Digitalisierung und Konvertierung der Dokumente, Metadaten- und Informationsdienstleistungen.

SHERPA (Securing a Hybrid Environment for Research Preservation and Access)¹⁸ ist ein dreijähriges von dem englischen Konsortium der Universitätsbibliotheken (CURL) finanziertes Projekt. Hierbei kooperieren sieben englische Universitäten¹⁹, die British Library und der Arts and Humanities Data Service mit dem Ziel, zahlreiche institutionelle elektronische Archive in England aufzubauen. *SHERPA* wendet die an der Universität Southampton entwickelte eprints.org.-Software an. Die Motivation für dieses Projekt ist die zunehmend kritische Haltung gegenüber den Rechteverwertern und ihrer wissenschaftsfeindlichen Preispolitik. Die Initiatoren von *SHERPA* streben zukünftig Verhandlungen mit Verlagen an, die es WissenschaftlerInnen ermöglichen sollen, ihre Arbeiten auch in einem nicht-kommerziellen Archiv zu veröffentlichen.

Bei *DARE* (Digital Academic Repositories)²⁰ handelt es sich um ein vierjähriges nationales Projekt (2003-2006) der niederländischen Universitäten unter Mitarbeit der Königlichen Bibliothek, der Königlichen Niederländischen Akademie der Wissenschaften und der niederländischen Organisation für wissenschaftliche Forschung. Ziel von *DARE* ist die Modernisierung der niederländischen Informationsinfrastruktur durch den Aufbau digitaler Repositorien, mit denen der wissenschaftliche „Output“ der Hochschulen elektronisch zugänglich und langfristig verfügbar gemacht werden soll. Unter Anwendung der internationalen Standards bilden die institutionellen Repositorien eine digitale Plattform für die akademische Information und sollen dazu beitragen, das Ansehen der niederländischen Hochschulen im internationalen Wettbewerb zu erhöhen. Die institutionellen Repositorien bilden die Grundlage für die Schaffung von Mehrwertdiensten, die durch die beteiligten Institutionen (selbst bzw. in Kooperation) in Zukunft entwickelt werden sollen.

Kurzfristig ist bei diesen Initiativen nicht davon auszugehen, dass sie die Veröffentlichung wissenschaftlicher Ergebnisse in renommierten Fachzeitschriften ersetzen. Fachzeitschriften werden auch im elektronischen Umfeld ihre Bedeu-

17 <http://www.dspace.org/>

18 <http://www.sherpa.ac.uk/documents/proposal.pdf>

19 University of Nottingham (Projektleitung), University of Edinburgh, University of Glasgow, Universities of Leeds, Sheffield and York, University of Oxford

20 <http://www.surf.nl/download/DARE-summary.pdf>

tung beibehalten. Nicht, weil sie *der Ort* wissenschaftlicher Kommunikation sind, sondern weil sie eine zentrale Rolle bei der Berufung und Beförderung von WissenschaftlerInnen spielen. Die Ziele, die sich mit dem Aufbau institutioneller Repositorien verbinden, liegen vielmehr in der Beschleunigung der Kommunikation der fachlichen communities untereinander, dem freien Zugang zur wissenschaftlichen Information, der Veröffentlichung von Forschungsergebnissen als preprints und als postprints, der langfristigen Sicherung der digitalen Dokumente und dem Renommee, welches die Institution durch den Aufbau der Repositorien in Zukunft erhalten kann.

4 Neue Kooperationen für die Publikation und Verbreitung wissenschaftlicher Information

Fachgesellschaften fehlt es häufig an den technischen und organisatorischen Kompetenzen und Voraussetzungen (Vertrieb, Marketing, Internettechnologie), um ihre Zeitschriften elektronisch verfügbar zu machen. Zwischen Fachgesellschaften, Bibliotheken, Universitäten, Universitätsverlagen und Marketingagenturen sind Kooperationsformen zu beobachten, durch die der Publikationsprozess im elektronischen Umfeld in geteilter Verantwortung neu organisiert wird.

*HighWirePress*²¹ ist ein Online-Verlag, der 1995 von der Stanford University Library gegründet wurde. Der Verlag arbeitet auf der Basis der Kostendeckung und verfügt derzeit über ein Zeitschriftenportfolio von etwa 300 Zeitschriften der Fachgebiete Biologie, Medizin und Lebenswissenschaften. Bei *HighWirePress* handelt es sich um einen Service Provider, der für die Fachgesellschaften eine elektronische Publikationsumgebung bereithält, damit diese ihre wissenschaftlichen Fachzeitschriften online präsentieren und distribuieren können. Die verlegerische Tätigkeit der Fachgesellschaften wird unterstützt, indem *HighWirePress* die technische Aufbereitung und Konvertierung in offene Dokumentformate, die Präsentation und Aufbereitung der Artikel und der Zeitschriften, die Beratung bei Geschäftsmodellen und das Hyperlinking und Crossreferencing der Dokumente übernimmt. Die von dem Verlag bedienten Fachgesellschaften wenden derzeit überwiegend das traditionelle Subskriptionsmodell in gekoppelter Form an. Darüber hinaus wird ein pay-per-view und ein pay-per-site angeboten. Der Verlag geht davon aus, dass sich die Kosten für die Produktion und Distribution ausschließlich elektronischer Dokumente nachhaltig reduzieren lässt. HighWire Press bietet ein großes Spektrum an archivierten Artikeln im Volltext an, die in eingeschränkter Form kostenfrei sind. Die Retrodigitali-

21 <http://highwire.stanford.edu/>

sierung vergangener Zeitschriftenexemplare soll in Zukunft mit Hilfe einer Stiftung finanziert werden.

Mit *ProjectMUSE*²² bietet der Universitätsverlag der Johns Hopkins University in Kooperation mit der Milton S. Eisenhower Bibliothek elektronische Ausgaben der in dem Verlag erscheinenden und von Fachgesellschaften herausgegebenen Fachzeitschriften aus den Fachgebieten der Sozial- und Geisteswissenschaften und der Mathematik an. Das Portfolio an Zeitschriften umfasst derzeit etwa 200 Zeitschriften. Der Universitätsverlag und auch das *ProjectMUSE* finanzieren sich zu 100% selbst. Als Publikationspartner der Fachgesellschaften übernimmt der Verlag das Marketing, die Websitegestaltung und den Vertrieb der angebotenen Zeitschriften. Das Geschäftsmodell des Verlages beruht auf dem traditionellen Subskriptionsmodell, jedoch werden die Zeitschriften zu wissenschaftsfreundlichen Preisen angeboten. Der Verlag operiert mit verschiedenen Varianten des Subskriptionsmodells und bietet unterschiedliche Pakete an.

Das im Jahr 2000 initiierte Projekt *BioOne*²³ ist eine Kooperation zwischen einem Online-Verlag, einer Marketing-Agentur, einem Bibliothekskonsortium mit 23 angeschlossenen Bibliotheken, Fachgesellschaften und der Universität von Kansas²⁴. *BioOne* produziert und distribuiert die Zeitschriften der Fachgesellschaften des American Institute of Biological Science. Die Unternehmung verfügt derzeit über ein Portfolio von 55 Zeitschriften, die überwiegend auf Subskriptionsbasis erhältlich sind. Um die Idee des kostenlosen Zugangs zu realisieren, werden neben dem traditionellen Subskriptionsmodell derzeit drei Zeitschriften frei zugänglich angeboten²⁵. Das kooperative Modell ermöglicht die preisgünstige Konvertierung der Printzeitschriften in elektronische Zeitschriften. Ist die Zeitschrift online, wird sie im Rahmen der *BioOne*-Sammlung an den Endnutzer weitergegeben. An der Finanzierung des Projektes *BioOne* ist ein Bibliothekskonsortium beteiligt, das nach Abschluss der Entwicklungsphase die Produkte von *BioOne* zu günstigeren Konditionen bezieht. Für die Verlage und Fachgesellschaften ist dadurch eine verlässliche Abonnentenbasis vorhanden. Weitere Ansätze zur Finanzierung der Online-Publikationen sind Werbung auf den Seiten des Online-Verlages, Sponsoring etc. *BioOne* bietet die Verlinkung der publizierten Artikel in die fachlichen Netzwerke an (z.B. BIOSIS;

22 <http://muse.jhu.edu/>

23 <http://www.bioone.org/bioone/?request=index-html>

24 vgl. hierzu Johnson, Richard K. (2000): A Question of Access. SPARC, BioOne, and Society-Driven Electronic Publishing. In: D-Lib Magazine, Vol.6, Nr.5
<http://www.dlib.org/dlib/may00/johnson/05johnson.html> (Zugriff 21.08.02)

25 vgl. ausführlicher zu weiteren Bestandteilen der Refinanzierung: Joseph, Heather D. (2003): BioOne: building a sustainable alternative publishing model for non-profit publishers. In: Learned Publishing, Vol. 16, S.134-138

NBII).²⁶ Für die Langzeitarchivierung der Dokumente kooperiert der Verlag mit PubMedCentral, die den freien Zugriff auf die Materialien gewährleistet. Darüber hinaus ist die Digitalisierung älterer Zeitschriftenexemplare geplant.

Eine Zusammenarbeit mit kleinen und mittelständischen Verlagen und Fachgesellschaften strebt *FIGARO* an. Es handelt sich hierbei um ein europäisches Projekt der niederländischen Universitäten Utrecht und Delft, der deutschen Universitätsbibliotheken Hamburg (Regionales Rechenzentrum) und Oldenburg, dem Verlag der Universität Florenz und dem Softwarehaus Daidalos in den Niederlanden.²⁷ Zielsetzung von *FIGARO* ist der Aufbau einer elektronischen Publikationsinfrastruktur auf einer europäischen Ebene, um Produkte der Verlage und Fachgesellschaften online anzubieten. Der modulare Aufbau von *FIGARO* soll die elektronische Publikation von Zeitschriften und Monographien, sowie die Publikation von preprints und postprints unterstützen. *FIGARO* strebt einen freien Zugang zu den digitalen Dokumenten auf der Basis des oben beschriebenen neuen Geschäftsmodells an.²⁸

5 Kostensenkung durch direkten Wettbewerb

Im Rahmen des „Alternative-Programms“ verfolgt *SPARC* die Strategie hochpreisigen, renommierten Zeitschriften direkte Konkurrenzprodukte als „head-to-head-competitor“ gegenüberzustellen²⁹ mit dem Ziel, die jährlichen Preissteigerungen bei den hochpreisigen Zeitschriften durch die Herausgabe eines deutlich preiswerteren Konkurrenzproduktes zu verlangsamen. Für die WissenschaftlerInnen bietet sich die Möglichkeit, durch ihr Publikationsverhalten aktiv auf die Preisentwicklungen großer kommerzieller Verlage Einfluss zu nehmen. Von *SPARC* werden diese Unternehmungen durch eine offensive Medienpräsenz unterstützt. Derzeit werden von den beteiligten Verlagen 14 alternative Zeitschriften in unterschiedlichen Fachgebieten herausgegeben. Das bekannteste Beispiel ist die 1999 herausgegebene *Evolutionary Ecology Research* als Konkurrenzprodukt zu *Evolutionary Ecology* von Wolters Kluwer. Das jüngste

26 vgl. Joseph, Heather: a.a.O., S.137

27 Daneben existieren weitere Kooperationspartner: Universitätsverlag Leuven, Forschungs- und Entwicklungsabteilung NetLab der Lund Universität, der Verlag Lemma in den Niederlanden und DiG in Polen, die Forschungsgruppe Delft Cluster und SPARC USA. Vgl. ausführlicher:

<http://www.figaro-europe.net/index2.html?mission.html>

28 Vgl. Savenije, Bas (2002): Who pays the Ferryman? About new models for scientific communication. Paper presented at the international symposium „Science & Engineering Libraries for the 21st Century“.

http://www.figaro-europe.net/docs/BAS_Leuven.doc

29 Algebraic and Geometric Topology, Documenta Mathematica, Journal of Machine Learning Research, Journal of Insect Science etc.

Beispiel ist die 2003 herausgegebene Zeitschrift „Journal of the European Economic Association“ als Konkurrent zu der von Reed Elsevier herausgegebenen *European Economic Review*. Einige dieser Zeitschriften sind bereits frei zugänglich, andere müssen durch die Bibliotheken subskribiert werden. Die Zeitschriften verfügen über Renommee und sind in einigen Fällen mit einem Gutachtergremium besetzt, welches zuvor für einen kommerziellen Rechteverwerter tätig war.

6 Bewertung der Initiativen und Unternehmungen

Der Zugang zur wissenschaftlichen Information bildet die Grundlage für die Produktion neuer Forschungsergebnisse. Die genannten Initiativen und Unternehmungen tragen auf unterschiedliche Weise dazu bei, die wissenschaftliche Informationsversorgung in einem digitalen Umfeld auf wissenschaftsfreundliche Weise neu zu gestalten. Dies, indem sie einen uneingeschränkten Zugriff auf die Information gewährleisten bzw. den Zugriff zu wissenschaftsfreundlichen Preisen ermöglichen. Insbesondere die institutionellen Repositorien und die auf freiem Zugang basierenden elektronischen Zeitschriften bieten derzeit das größte Potenzial zur Reformierung des wissenschaftlichen Publikations- und Kommunikationsprozesses. Zugleich eröffnen sich flexiblere Kommunikations- und Publikationsmöglichkeiten und die Möglichkeit den Begutachtungsprozess im elektronischen Umfeld transparenter zu gestalten. Mittelfristig kann dies zu einer Stärkung der Position der Hochschulen und außeruniversitären Forschungseinrichtungen im internationalen Wettbewerb beitragen und eine strukturelle Erneuerung des wissenschaftlichen Publikationsprozesses befördern, die auch solchen Ländern Partizipationschancen bietet, denen dies aufgrund der Kosten für wissenschaftliche Information bislang nur eingeschränkt möglich ist.

Qualitätssicherung und Nutzung von wissenschaftsrelevanten Angeboten auf dem Deutschen Bildungsserver: Empirische Befunde einer vergleichenden Logfile-Analyse

Elke Brenstein, Olaf Kos

Abteilung für Pädagogik und Informatik
Humboldt Universität zu Berlin

Wie können wir mehr Menschen mit mehr Informationen versorgen, schnell und in verschiedenen Formen? Forscher haben hundert Jahre an dem Problem gearbeitet.

Und die wunderbare Nachricht ist, dass sie es tatsächlich gelöst haben – bravo!

Nur ist dabei ein neues Problem entstanden: Jetzt wissen die Menschen nicht, was sie mit all den Informationen anfangen sollen....

Neil Postman, 1996

Abstract

Informationen im Web werden in zunehmendem Maße thematisch gebündelt und über Internetportale für eine heterogene Nutzerschaft zugänglich gemacht. Auch im Bereich der Wissenschaften können Portale den Wissensaustausch vereinfachen, indem sie die Orientierung und den Umgang mit Fachinformationen erleichtern. Inwieweit Portale die Informations- und Kommunikationsbedürfnisse von Wissenschaftlern unterstützen, hängt maßgeblich von der Qualität der Inhalte und deren Aufbereitung ab. In unserem Beitrag werden allgemeine Qualitätskriterien für Internetportale in der Wissenschaft vorgestellt und empirische Daten zur Nutzung von wissenschaftsrelevanten Ressourcen beim Deutschen Bildungsserver vorgestellt. Die Ergebnisse einer vergleichenden Logfile Analyse geben Auskunft über die Nutzungsentwicklung der Angebote und zeigen auf, welche Inhalte beim Suchen und Browsen in den Bereichen Hochschule, Wissenschaft und Forschung besonders nachgefragt werden. Aus den Ergebnissen werden Empfehlungen zur nutzergruppenspezifischen Optimierung der Angebote abgeleitet.

Portale als Wegweiser im Informationsdschungel

Seit dem Durchbruch des WWW als Massenmedium in der 2. Hälfte der 90er Jahre hat sich das Internet sowohl in seinen technischen Möglichkeiten als auch

mit seinen Inhalten explosionsartig entwickelt. Diese Informationsexplosion erzeugt eine Vielzahl von Problemen:

- So ist der vereinfachte Zugriff auf Informationen häufig mit dem Verlust an Orientierung bei der Informationsgewinnung verbunden. Die unübersichtliche Quellenlage erfordert zudem eine höhere Wertungs- und Selektionskompetenz auf Seiten der Nutzer.
- Der hohen Quantität an Informationen im Internet steht oft eine vergleichsweise geringe Zahl qualitativ hochwertiger Informations- und Kommunikationsangebote gegenüber.
- Mit der exponentiellen Zunahme an Informationen sinkt deren prozentualer Anteil, der über Suchmaschinen gefunden werden kann; trotz zunehmender Leistungsfähigkeit der Maschinen.

Suchmaschinen fördern zu viele, oft wenig relevante Informationen zu Tage. Der Nutzer bekommt dabei meist nur zu sehen, was an der Oberfläche von Webseiten zu finden ist. Wie beim Auswerfen eines Fischernetzes bleibt vieles, was sich in den Tiefen dynamischer Datenbanken verbirgt, für den Nutzer unauffindbar. Noch gibt es keine allgemein verfügbaren Lösungen, welche die Inhalte, die in den Tiefen und Weiten des Internets versteckt liegen, mit hoher Zielgenauigkeit recherchierbar machen. Das Medium Internet wird unübersichtlicher und seine Informationseffizienz für seine Nutzer unbefriedigender.

Eine Antwort auf die Frustration der Nutzer bei der Informationssuche im Netz sind übersichtlich strukturierte und redaktionell gepflegte Themenangebote in Form von Internetportalen. Daher steigt die Anzahl zentraler Einstiegsseiten, die es sich zur Aufgabe gemacht haben, eine Orientierungshilfe mit informationellem Mehrwert zur Verfügung zu stellen, stetig an.

Sowohl für kommerzielle Portale, welche aus finanziellem Interesse um die Gunst der Nutzer buhlen, als auch für non-profit Angebote im Bildungsbereich gilt, dass immer höhere Qualitätsmaßstäbe angelegt werden müssen. So haben sich die Standards hinsichtlich der Auswahl und Darstellung der Inhalte und der Funktionalität von Webseiten, welche in den Anfängen auf gut strukturierte Linklisten beschränkt waren, drastisch verändert.

Angesichts der wachsenden Anzahl von Portalangeboten im Bildungsbereich ist die *Qualität der Angebote* von entscheidender Bedeutung für die Effizienz und Effektivität von Informationsgewinnung und -verarbeitung im Internet. Eine Antwort auf die Frage nach der Definition von Qualität ist nicht einfach, da es viele Faktoren und Sichtweisen zu berücksichtigen gilt. Da ist zum einen die Sicht derer, die ein Informationsangebot aufbereiten, aus technischer, graphischer und inhaltlicher Sicht. Ähnlich wie bei der Metapher von den drei Blinden und dem Elefant kann man auf die Frage „Was macht die Qualität eines Internetportals aus?“ durchaus unterschiedliche Antworten von einem Webde-

signier, einem Programmierer oder einer Redakteurin erhalten, da jede/r unterschiedliche Schwerpunkte bei der Bewertung von Qualität setzt.

Interessanterweise klaffen auch die Bewertungen von Anbietern und Anwendern oft weit auseinander. Grafische Elemente, die für den Webdesigner hohen ästhetischen Wert haben, können vom Nutzer einfach als störend bei der Suche nach Informationen wirken. Wesentlich für den Erfolg eines Internet Angebots ist daher die Bewertung durch die Nutzer (sowie auch die Einschätzungen von Nichtnutzern). Aus diesem Grund ist eine empirische Untersuchung der Nutzung und der nutzergruppenspezifischen Bewertungen des Informationsangebotes für eine marktgerechte und zukunftssträchtige Entwicklung von Bildungsportalen unerlässlich.

Im folgenden wird zunächst der Begriff der Qualität aus Expertensicht definiert. Anschließend werden Ergebnisse einer empirischen, zielgruppenspezifischen Untersuchung der Nutzung des Deutschen Bildungsservers vorgestellt. Daraus werden Empfehlungen für die nutzergruppenspezifische Optimierung der Angebote abgeleitet.

Qualitätskriterien für Internetportale

Portale werden u.a. unterschieden in *inhaltsorientierte* (horizontale) Portale, welche Inhalte für bestimmte Interessengemeinschaften vorhalten und *zielgruppenorientierte* Portale, welche ein Angebotsmix für bestimmte Gruppen anbieten¹. Bei den meisten Bildungsportalen werden Informationen *sowohl* themenorientiert (z.B. Schule, Hochschule, Weiterbildung) *als auch* zielgruppenorientiert (z.B. Lehrer, Schüler, Eltern, Studenten) angeboten. Aus der Heterogenität der Inhalte und der Nutzergruppen ergeben sich somit besondere Herausforderungen im Hinblick auf (1) die Gestaltung der Seiten, d.h. Layout und Design, (2) Aufbau und Struktur der Inhalte (3) die technische Funktionalität und (4) besondere Serviceangebote.

1 Gestaltung der Seiten

Eine Portalseite sollte auf den ersten Blick als solche zu erkennen sein und dem Benutzer ein übersichtlich gestaltetes Angebot präsentieren.

„Corporate Identity“. Ein Portal zeichnet sich nicht nur durch eine „single-point-of-access“ Bündelung von Inhalten und Service Angeboten aus, sondern vor allem durch eine einheitliche Benutzeroberfläche, mit einheitlicher Darstellung von Inhalten und Navigationselementen. Auch sollten Farbschemata und graphische Elemente innerhalb der Site konsistent verwendet werden. Das Gefühl der Corporate Identity ergibt sich aus der Einheit von identitätsstiftenden

1 http://www.teachsam.de/arb/internet/WWW/arb_www_portale_2.htm

Symbolen, wie Logo, Schriftarten und –farben und Symbolfamilien. Alles in allem sollte ein Portal ein harmonisches Bild abgeben, welches mit anderen Produkten der Öffentlichkeitsarbeit (Broschüren, Briefpapier) etc. übereinstimmt.

Übersichtlichkeit. Von der graphischen Gestaltung abgesehen, sollte das Angebot vor allem übersichtlich sein. Das bedeutet, dass sowohl für den unerfahrenen Nutzer mit wenig inhaltlichem Vorwissen klar erkennbar sein sollte, welche Arten von Informationen wo zu finden sind, als auch dem Experten möglichst schnell Zugang zu speziellen Themen ermöglicht werden soll. Besonders wichtig ist, dass einzelne Funktionsbereiche wie Suche, Aktuelles, Themenkataloge sowie nähere Informationen über die Site bzw. deren Betreiber (Kontakt, Hilfe, Impressum, Sitemap etc.) leicht als solche identifiziert werden können. Oft wird das Auffinden von wichtigen Informationen durch ungewöhnliche Platzierung oder Linkbenennungen erschwert. Bei der Verwendung von graphischen Elementen ist eine Unterlegung mit einem erklärenden Text, der beim Zeigen mit der Maus erscheint, hilfreich.

Navigation. Die Navigation innerhalb und zwischen den Bereichen sollte einfach und selbsterklärend sein und, wo möglich, durch Navigationshilfen unterstützt werden. Außer den gängigen hierarchischen Menüleisten können bei komplexen Auswahloptionen Auswahllisten eingesetzt werden. Ganz wichtig ist die Bereitstellung einer Tiefennavigation (Pfad- oder bread-crumb Navigation), die dem Benutzer die Orientierung in einer hierarchischen Seitenstruktur erleichtert. Der Nutzer sollte immer in der Lage sein, zu wissen wo er/sie herkommt, bzw. klar erkennen können, wo es hin geht. Dafür können als weitere Navigationshilfen *Site Indizes* oder *Sitemaps* zur Verfügung stehen.

Gestaltungsstandards. Inzwischen haben sich gewisse Standardlösungen für die Aufteilung von Seiten und die Platzierung von Navigationselementen etabliert. Es ist daher sinnvoll, sich an existierende Gestaltungsstandards anzulehnen, um die „Lernkurve“ beim Erstbesuch so gering wie möglich zu halten (Nielsen, 2001; Krug, 2002). Ungewöhnliche Icons oder Navigationsstrukturen können aus künstlerischer Sicht ihren Reiz haben, können bei informationsorientierten Seiten jedoch manchmal wenig hilfreich sein. So kann z.B. die Verwendung von Roll-out Menüs (JavaScript) im Navigationsbereich problematisch sein, da Nutzer die Möglichkeiten nur sehen, wenn sie den Bereich mit der Maus berühren.

2 Aufbereitung der Inhalte

Obwohl bei Erstbesuchern der erste Eindruck darüber entscheidet, ob sie sich ein neues Angebot näher anschauen oder zum nächsten Ergebnis der Suchmaschinenliste übergehen, zeigt sich tendenziell, dass die graphische Aufbereitung

im Bildungsbereich einen wesentlich geringeren Stellenwert hat, als die Menge und Qualität der Inhalte (Brenstein & Kos, 2004).

Thematische Vielfalt. Ein gutes Portal zeichnet sich daher in erster Linie sowohl durch die Breite als auch die Tiefe des Informationsangebotes aus. Dabei sind über die Versorgung mit Informationen und Ressourcen zu allgemeinen Themen hinaus, besondere Themen und Angebote ein wichtiger Wettbewerbsvorteil. Diese gilt es als Marktnischen zu identifizieren und auszubauen, möglicherweise über Kooperationen mit anderen Anbietern.

Angebote für verschiedene Zielgruppen. Nutzer suchen in erster Linie Informationen, die für sie nützlich sind. Daher ist es sinnvoll, nutzergruppenspezifische Informationen und Zugänge zu Querschnittsthemen zu realisieren. Je mehr relevante Informationen gut gegliedert vorhanden sind, desto größer sind die Chancen, dass Nutzer verweilen und wiederkommen.

Aktualität und Seriosität. Natürlich ist nicht allein die Menge entscheidend sondern auch die Seriosität der Information. Besonders in Anbetracht vieler redundanter Angebote zu manchen Bildungsinformationen ist es wichtig zu wissen, wer der Autor ist, bzw. für das Angebot verantwortlich ist. Maßgeblich für eine Einschätzung der Qualität von Internetinformationen ist auch die Angabe, wann ein Beitrag oder eine Linkliste erstellt, bzw. zuletzt aktualisiert wurde. In vielen Fällen sucht man vergeblich nach Quellen- und Datumsangaben. Internetseiten sollten daher vom Betreiber außerdem regelmäßig anhand einer automatischen Linküberprüfung auf ihre Aktualität überprüft werden, um zu vermeiden, dass Benutzer auf „tote Links“ stoßen.

Hilfe. Die Benutzerfreundlichkeit von Portalen kann durch die Bereitstellung von allgemeinen Informationen zur Orientierung in der Site sowie Erklärungen zu besonderen Funktionen (z.B. Suchfunktion) maßgeblich verbessert werden. Auch können *Guided Tours* Nutzer mit den wichtigsten Portalangeboten vertraut machen. Eine äußerst effektive Form Hilfestellung zu leisten ist das Pflegen einer *FAQ-Liste*, in der die häufigsten Fragen und Problemlösungen dokumentiert sind. Damit ersparen sich Benutzer langwieriges Suchen nach oder Warten auf eine Antwort. Gleichzeitig entlastet eine FAQ-Liste Mitarbeiter, die Nutzeranfragen beantworten. Weiterhin kann ein *Glossar* besonders für Novizen eine wichtige Hilfestellung zum Verständnis der Inhalte sein.

3 Technische Funktionalität

Ladezeiten. Der motivierende Effekt der von interessanten Inhalten und/oder einer ansprechenden grafischen Gestaltung ausgeht, kann leicht zunichte gemacht werden, wenn die Seiten nicht wie erwartet funktionieren. Das kann verschiedene Ursachen haben. Manchmal sind grafische Elemente nicht optimiert und führen zu langen Ladezeiten. In einem schnelllebigen Medium wie dem Internet

mit vielfältigen Alternativangeboten, springen Benutzer leicht ab, wenn Seiten zu lange laden.

Browserdarstellung. Webseiten sollten auf validiertem HTML 4.01 und CSS nach den offiziellen W3C-Standards basieren, um eine einheitliche und fehlerfreie Darstellung auf allen modernen Browsern zu gewährleisten. Die Kompatibilität kann mit dem W3C HTML-Validator² kostenlos überprüft werden. Webseiten sollten nicht für veraltete Browserversionen (z.B. Netscape 4.x) optimiert werden, da sie dann nicht vorwärtskompatibel und zukunftsfähig sind.³ Seiten sollten zudem so programmiert sein, dass sie auch ohne Plug-Ins und andere technische Besonderheiten funktionieren.

Auflösung. Obwohl sich Standardgrößen und –auflösungen für Computermonitore ständig vergrößern, ist es wichtig zu berücksichtigen, dass im Bildungsbereich noch viele Nutzer ältere Systeme verwenden und daher Webseiten sowohl für eine Auflösung von 600x800 also auch für höhere Auflösungen programmiert werden sollten.

Suchfunktion. Von zentraler Bedeutung für die inhaltliche Erschließung einer Website ist die Suchfunktion. Studien zeigen wiederholt, dass die meisten Nutzer zielgerichtet nach Informationen anhand von Schlagworten suchen anstatt sich durch Themenhierarchien zu klicken (Nielsen 2001). Idealerweise sollte sowohl eine Schnellsuche als auch eine Erweiterte Suche mit Volltextrecherche-Möglichkeiten vorhanden sein. Wenn möglich, sollte die Ausgabe der Suchlisten nach einzelnen Datenfeldern variierbar sein (z.B., Datum, Autor, Schulfach).

Druckfunktion. Da Inhalte im Web meist in seitengroße, hypermedial vernetzte Inhaltseinheiten aufgeteilt sind, ist es oft mühsam einen größeren, zusammenhängenden Block zum Offline Lesen und Archivieren auszudrucken. Auch sind die Ausdrücke von Seiten, mit vielen farbigen Bildern oder Hintergrundelementen oft schwer lesbar und verbrauchsmittelintensiv (Papier und Toner). Daher ist es wünschenswert, dass wichtige Inhalte nicht nur als hypermediale Webseiten sondern auch in einer linearen Textversion vorliegen, bzw. benutzerfreundlich ausdrückbar sind.

Barrierefreiheit. Webseiten sollten nach den Zugänglichkeitsrichtlinien der W3C Web Accessibility Initiative (WAI) behindertengerecht gestaltet werden⁴. Diese sind in drei Prioritätsstufen gegliedert. So sollten z.B. immer alternative Darstellungsformate (Text statt Bild, Audio zusätzlich zu Text oder Bild) zur Verfügung gestellt werden, einfache Farbkontraste gewählt werden, die Darstellung ohne Stylesheets ermöglicht werden. Auf jeden Fall sollte die Schriftgröße variabel sein.

2 <http://validator.w3.org/>

3 <http://www.scoreberlin.de/accessibility.shtml>

4 <http://www.w3.org/Consortium/Offices/Germany/Trans/WAI/webinhalt.html>

4 Serviceangebote und Kommunikationsdienste

Newsletter. Da die meisten Nutzer viele Websites nicht regelmäßig besuchen, bietet es sich an, interessierte Nutzer in Form eines Newsletters über neue Entwicklungen auf der Website (z. B. Personalien, neue Themen, Bereiche oder Ressourcen) sowie über aktuelle Themen (Veranstaltungen, Neuigkeiten) aus den jeweiligen Inhaltsbereichen zu informieren.

Personalisierung. Immer mehr Portalanbieter ermöglichen einen personalisierten Zugang zu ihren Angeboten. Meist werden dabei Nutzerdaten mittels „cookies“ gespeichert und weitergereicht. Wer diese Funktionalität im Browser aus Sicherheitsgründen abgeschaltet hat, kann diese Angebote nicht oder nur bedingt nutzen. Bei einem persönlichen Zugang wird dem Nutzer ermöglicht, ein individuelles Nutzerprofil zu erstellen und gezielt und interessengeleitet informiert zu werden, entweder auf den Portalseiten selbst oder in Form eines individuell konfigurierten Newsletters. In manchen Fällen lassen sich auch Darstellungsoptionen auswählen.

Kommunikation. Der persönliche Austausch zwischen Nutzern mit gemeinsamen Interessen kann sehr produktiv sein. Wenn sich Interessengemeinschaften zu einzelnen Themen bilden lassen, dient es der Kundenbindung, wenn Nutzer über eine Portalseite an einem *Online Forum* oder einer *Mailingliste* teilnehmen können. Besonders bei jungen Nutzern ist die offene synchrone Kommunikation im *Chat* sehr beliebt. Aber auch andere Nutzergruppen finden sich zu themenzentrierten Chats zusammen. Diese funktionieren am besten, wenn sie moderiert werden. Noch weniger üblich, da technisch aufwändig sind *Video-konferenzen*. Diese Möglichkeiten werden mit zunehmender Bandbreite und Verfügbarkeit von kostengünstigen technischen Lösungen an Bedeutung gewinnen und den Austausch in Interessengemeinschaften beflügeln. Eine weitere Form der Kommunikation zwischen Nutzern ist das *Weiterleiten oder Empfehlen von interessanten Seiten*, welches von manchen Portalanbietern in Form eines Formulars angeboten wird. Schließlich gewinnt ein Portalangebot an Lebendigkeit, wenn den Nutzern ermöglicht wird, ihre Meinungen in Form von *Schnellabstimmungen* (deren Resultate umgehend für andere zu sehen sind) zu äußern. Diese werden häufig zur Erfassung von Trends zu aktuellen Themen genutzt.

Empirische Befunde zur Nutzung des Deutschen Bildungsservers

Der Deutsche Bildungsserver ist ein Meta-Server, welcher seit 1996 bildungsrelevante Informationen für eine breite Palette von Nutzern zur Verfügung stellt. Um den Anforderungen einer wachsenden Nutzerschaft (siehe Abb. 1) gerecht zu werden, wurde im Jahre 2000 ein Redesign durchgeführt, bei dem die o.g. Qualitätskriterien soweit wie möglich berücksichtigt wurden.

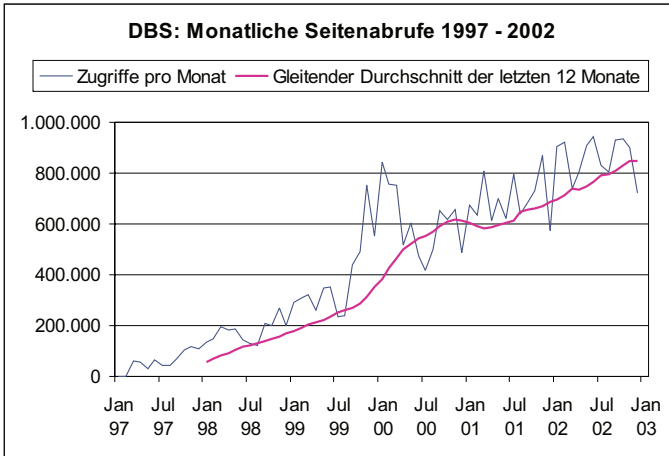


Abbildung 1

Steigende Zugriffszahlen⁵ und informelles Nutzerfeedback belegten, dass das Angebot von den Adressaten in den Bereichen Schule, Hochschule, Weiterbildung und Erwachsenenbildung, Berufliche Bildung und Bildungsadministration im Allgemeinen gut angenommen wurden. Bisher war jedoch nur unzureichend bekannt, wie sich die Nutzung auf unterschiedliche Gruppen verteilt und inwieweit spezifische Nutzergruppen mit den Angeboten und ihrer Präsentation zufrieden waren. Eine Vorstudie der Abt. Pädagogik und Informatik der HU Berlin (Brenstein & Kos, 2004) hat erste Anhaltspunkte zur Verteilung der Nutzergruppen, ihres Nutzungsverhaltens und ihrer Zufriedenheit mit den Inhalten und der Aufbereitung und Gebrauchstauglichkeit der Serverangebote ergeben. Dabei ergaben sich unterschiedliche, sich ergänzende Befunde aus den verschiedenen methodischen Untersuchungsansätzen (Online Befragung, Logfile-Analysen und Usability Untersuchungen).

Die Ergebnisse dieser ersten nicht-repräsentativen Nutzerstudie (N=823) zeigten, dass die Angebote des Deutschen Bildungsservers von Vertretern beider Geschlechter und aller Altersgruppen gleichermaßen genutzt wurden. Dabei bildeten Lehrer und Lehrerinnen die stärkste Gruppe, gefolgt von Wissenschaftlern. Die meisten Besucher wurden durch einen Link auf einer anderen Webseite oder durch einen Hinweis von einem Kollegen oder Bekannten auf den Deutschen Bildungsserver aufmerksam. Sie suchten hauptsächlich nach Lehr-/Lernmaterialien und -ressourcen. Die am häufigsten aufgerufenen Seiten bezogen sich direkt oder indirekt auf den Bereich Schule.

⁵ Die Anzahl der monatlichen Seitenabrufe basiert auf Logfile-Statistiken erhoben durch `http-analyse 2.4`.

Insgesamt betonten die Nutzer die hohe Relevanz derartiger strukturierter Internetangebote für erfolgreiches Wissensmanagement und lobten sowohl die Vielfalt als auch die Qualität der Inhalte. Jedoch gab es auch zahlreiche Vorschläge zu inhaltlichen Ergänzungen und Verbesserungen der Angebote. Diese betraf vor allem die Strukturierung und Darstellung der Angebote sowie Aspekte der Benutzerführung und Funktionalität, wie z.B. die Suchfunktion. Besonders informativ waren die detaillierten Kommentare, die sich aus der intensiven Beschäftigung mit Inhalten, Layout & Design, Gebrauchstauglichkeit und Funktionalität in mehreren Usability-Untersuchungen mit Teilnehmern aus den Bereichen Schule, Hochschule und Weiterbildung (N= 101) im Rahmen der o.g. Vorstudie (Brenstein & Kos, 2004) ergeben hatten. Aus vielen Kritikpunkten konnten direkte Handlungsanweisungen abgeleitet werden, wie zum Beispiel die Reduzierung der Komplexität der Informationen auf der Homepage. Jedoch wurde bei manchen Themen deutlich, dass nur eine Befragung einer genügend großen repräsentativen Stichprobe aller relevanter Nutzergruppen valide Antworten auf strittige Fragen liefern kann. Im Zuge der Qualitätsentwicklung und -sicherung der Angebote des Deutschen Bildungsservers läuft derzeit eine repräsentative Untersuchung in den Bereichen Hochschule, Schule und Weiterbildung zur Erfassung von Nutzungsintentionen und Meinungen der jeweiligen Nutzergruppen⁶

Vergleichende Logfile Analyse

Befragungen, ob postalisch oder online durchgeführt, vermitteln jedoch nur einen begrenzten Einblick in das Nutzerverhalten, da nur Antworten von einem Teil der Nutzer vorliegen. Selbst wenn bei der Befragung die Repräsentativität gewährleistet ist, bleiben die Befunde nur bedingt aussagekräftig. Das Gleiche gilt für Usability Studien, welche wertvolle Detailinformationen bieten, aber nichts über die Gesamtnutzung aussagen. Daher wurden in der DBS Nutzerstudie zunächst die Server Logfile-Daten für den Zeitraum eines Jahres analysiert, um Nutzungshäufigkeiten, Informationen über Hosts (Domänen & Organisationen), verwendete Browser und Betriebssysteme, sowie Referrer und am häufigsten besuchte Seiten und Suchworte zu erhalten.

Identifizierung einer Hochschul- und Forschungsstichprobe

Ein Problem bei der Analyse von Logfile Daten ist, dass zwar allgemeine Aussagen über Nutzungsmuster gemacht werden können, aber kaum nutzergruppenspezifische Aussagen getroffen werden können, da aus den Logfile-Daten in der Regel nicht ersichtlich ist, wer welche Angebote genutzt hat (Nutzeridentifizierung über „cookies“ wird beim Deutschen Bildungsserver bewusst vermie-

6 <http://www.bildungsserver.de/zeigen.html?seite=1541>

den). Allgemein ist es schwierig, die Anzahl und Herkunft der Besucher genau zu bestimmen, da dynamische Internetadressen oder Proxies dies verhindern können. Jedoch kann, zum Beispiel, die Hostadresse einen Hinweis auf die „Herkunft“ der Zugriffe geben und als Heuristik für die Zielgruppenbestimmung verwendet werden.

Da die Mehrzahl der Zugriffe von Nutzern aus dem Bereich Hochschule und Forschung über Adressen von universitären oder außeruniversitären Einrichtungen erfolgt, war es möglich, die Zugriffe dieser Zielgruppe annähernd zu identifizieren. Dafür wurde auf der Grundlage einer Liste aller Universitäten und Forschungsinstitute eine Teilstichprobe aus der Gesamtmenge der Nutzer (Sept. 01 bis Okt. 02) ermittelt und definiert. Die Anfragen der entsprechenden Adressen wurden aus der Gesamtmenge der Logfiles heraus gefiltert und getrennt analysiert. Gefundene Unterschiede zwischen der Teilstichprobe von Nutzern aus dem Bereich Hochschule und Forschung (H&F Stichprobe) und der Gesamtmenge sollten auf besondere Eigenschaften dieser Gruppe von Benutzern zurück zuführen sein.

Bevor die Daten untersucht werden konnten, war es notwendig, Zugriffe von Robotern und Webcrawlern zu identifizieren und in einem iterativen Verfahren zu eliminieren. Obwohl automatische Programme einen Großteil der Roboterzugriffe heraus filtern, bleiben doch oft einige Roboter unerkannt. Zur Erstellung benutzerdefinierter Analysen und detaillierter Gesamtberichte wurden die kostenlos verfügbaren Logfile Analyse Tools Analog und *Report Magic*⁷ verwendet (Analog ermöglicht das individuelle Editieren einer Konfigurationsdatei, welche die Ausführung der Logfile-Analysen steuert). Darüber hinaus wurden Skripte geschrieben, um die Programmierung spezifischer Analysen zu vereinfachen. Es wurden insgesamt 7,700,473 Seitenabrufe analysiert.

Ausgewählte Ergebnisse

Die in der H&F Stichprobe enthaltenen Internetadressen machten nur 4% der gesamten Anzahl der Anbieter aus. Entsprechend entfielen auch nur 5% der aufgerufenen Daten auf die H&F Stichprobe. Auch der Anteil an Seitenanfragen war ähnlich gering (6% H&F) (siehe Abb. 2). Wie aus Abb. 3 ersichtlich ist, wurden in der H&F Stichprobe jedoch proportional wesentlich mehr verschiedene Seiten aufgerufen (17%). Somit scheint es, dass Nutzer aus dem Wissenschaftsbereich die Seiten des Deutschen Bildungsservers für differenziertere Recherchen nutzen als die Gesamtnutzerschaft.

Internet Anbieter. In der Gesamtstichprobe wurden etwa ein Drittel (34 %) der Anfragen durch den Internet Service Provider „t-dialin.net“ getätigt. Daraus ist ersichtlich, dass ein Großteil der Nutzer über Modem auf den Bildungsserver

7 <http://www.reportmagic.org/>

zugreift. Dies deckt sich mit den Befunden aus anderen Analysen nach denen die Hauptnutzergruppe aus Lehrern und Lehrerinnen besteht, die auch von zuhause auf den Bildungsserver zugreifen. Weitere ISPs haben einen jeweils deutlich kleineren Anteil an den Anfragen: aol.com (7 %), pppool.de (5 %), belwue.de (4 %) und t-online.de (3 %). Die Detailanalyse der Hostadressen der H&F Stichprobe ergab, dass die meisten Anfragen von den Universitäten Bielefeld und Dortmund gestellt wurden.

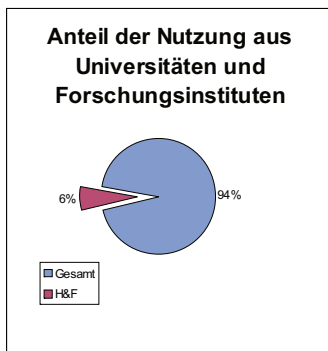


Abbildung 2

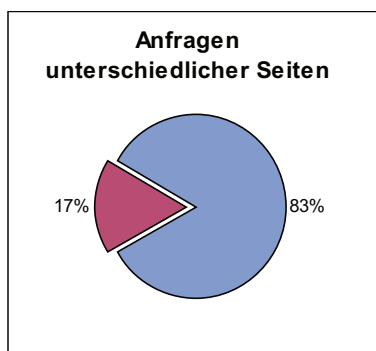


Abbildung 3

Betriebssysteme und Browser. Ungefähr 90% aller Seitenanfragen wurden von Computern getätigt, die Windows als Betriebssystem installiert haben. Der Anteil von Macintosh ist in beiden Stichproben mit 2% sehr gering. Erwartungsgemäß war der Anteil von Unixnutzern in der H&F Stichprobe etwas größer als in der Gesamtstichprobe (5% versus 1%).

Wie aus Abb. 4 ersichtlich ist, war der am häufigsten benutzte Webbrowser in beiden Stichproben der Internet Explorer (75% IE 5, 25% IE 6), wobei der Anteil in der H&F Stichprobe *deutlich* geringer war als in der Gesamtstichprobe (51% versus 82%). In der H&F Stichprobe wird interessanterweise Netscape viel häufiger benutzt (48% versus 15%).

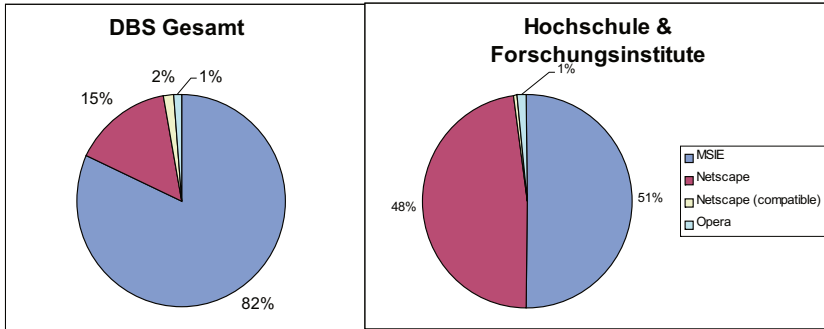


Abbildung 4

Referrer. Aus dem Referrer-Eintrag ist ersichtlich, von welcher Internetseite aus eine Anfrage gemacht wurde (wenn die Seite über ein Lesezeichen aufgerufen, oder die Seitenadresse in der Adressenleiste eingegeben wird, wird kein Referrer registriert). Die Analyse der externen Referrer (Seitenanfragen, über Links auf Webseiten, die nicht dem Bildungsserver selbst angehören) ergab, dass in beiden Stichproben die Mehrzahl der Zugriffe über Seiten von Suchmaschinen (hauptsächlich Google) erfolgte (59 % in der Gesamtstichprobe, 59% in der H&F Stichprobe). Die Analyse der restlichen Sites ergab, dass in der H&F Stichprobe mehr Besucher von Seiten kommen, die von Universitäten und Instituten betreut werden als in der Gesamtstichprobe (35% versus 13%) und 5% von studieren-im-netz.de auf den Deutschen Bildungsserver verwiesen werden. Dagegen kommen in der Gesamtstichprobe etwa doppelt so viele Besucher von Schule.de Seiten als in der H&F Stichprobe (32 % versus 14 %). Auch überwiegt der Anteil von Verweisen von Landesbildungsservern (13 % versus 5 %) und dem SchulWeb (18 % versus 12 %) in der Gesamtstichprobe.

Suchanfragen. Wenn jemand mit Hilfe einer Suchmaschine zum Bildungsserver gelangt, wird auch der eingegebene Suchbegriff in den Logfiles registriert. Für beide Stichproben wurden daher die 30 häufigsten Suchanfragen analysiert. Diese wurden thematisch in Kategorien gruppiert und der relative Anteil dieser Gruppen von Suchbegriffen verglichen. Die Top 30 Suchanfragen machten in beiden Gruppen etwa ein Viertel aller Suchanfragen aus. Wie aus Abb. 5 ersichtlich ist, wurde in beiden Stichproben am häufigsten nach Begriffen zum Thema PISA-Studie gesucht, gefolgt von der Kategorie Bildungsserver. In der H&F Stichprobe wurde überdurchschnittlich häufig nach Begriffen in der Kategorie Hochschule gesucht. Vergleichsweise wurde weniger häufig nach Begriffen aus den Bereichen Schule und Weiterbildung gesucht.

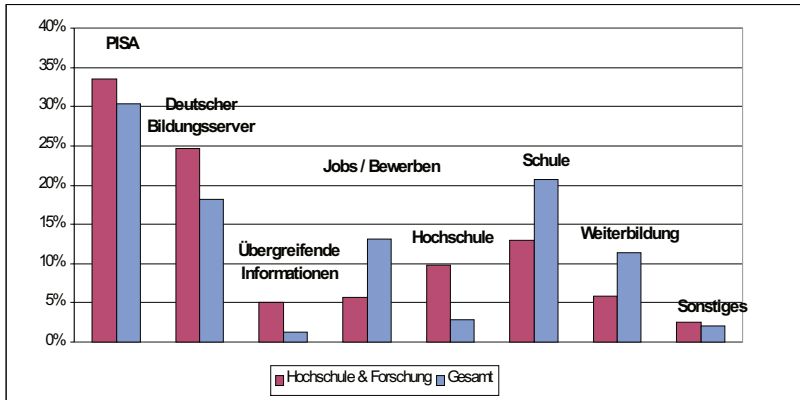


Abbildung 5

Seitenabrufe in einzelnen Bereichen

Die Analyse der Suchworte hat deutlich gezeigt, dass sich Nutzer aus Hochschule und Forschung für Inhalte aus dem Bereich Hochschule interessieren, wenn sie die Suchfunktion des Bildungsservers nutzen. Bei der allgemeinen Fragestellung nach nutzergruppenspezifischen Zugriffen war weiterhin von Interesse, inwieweit a) allgemeine Informationsseiten, b) Informationen zu einzelnen Bildungsbereichen und c) Angebote für einzelne Nutzergruppen unterschiedlich genutzt werden.

Die Seiten mit *allgemeinen Informationen* zum Deutschen Bildungsserver wurden insgesamt nur wenig genutzt (weniger als 1 % der Seitenzugriffe entfielen auf diesen Bereich). Dabei ergab die vergleichende Analyse der Zugriffe, dass kaum nennenswerte Unterschiede zwischen den Stichproben bestehen, mit Ausnahme der Kontaktseiten, die von Nutzern aus dem Bereich Hochschule und Forschung deutlich weniger genutzt werden.

Wie aus Abb. 6 ersichtlich ist, waren die Zugriffe auf Seiten mit *Inhalten zu den einzelnen Bildungsbereichen* in beiden Gruppen bedeutend höher (bis zu einem Viertel aller Seitenzugriffe), vor allem im Bereich *Übergreifende Informationen* und *Schule*. Wider Erwarten werden die für Hochschule und Forschung relevanten Seiten nicht in erster Linie von ihrer Zielgruppe frequentiert. Die Bereiche Hochschulbildung (Gesamt 6%, H&F 7%) und Wissenschaft und Bildungsforschung (Gesamt 1,3%, H&F 1,6%) werden von allen Nutzern wenig genutzt. Interessant ist, dass auch die Seiten im Bereich Schule vergleichsweise häufig von Nutzern aus Hochschule und Forschung genutzt werden (Gesamt 27%, H&F 22%). Dies mag damit zusammen hängen, dass sich viele Wissenschaftler thematisch mit schulrelevanten Themen auseinandersetzen oder mög-

licherweise auch in ihrer Rolle als Eltern Informationen zum Thema Schule recherchieren.

Ein klarer Unterschied zwischen den beiden Stichproben war jedoch bei den nutzergruppenspezifischen Zugängen in der Rubrik „Angebote Für“ zu verzeichnen (Abb. 7). Obwohl diese Seiten mit bis zu 10% aller Zugriffe insgesamt weniger häufig frequentiert wurden, wurden Seiten mit spezifischen Informationen für Studierende (Gesamt 4,9%, H&F 9,5%) und Bildungswissenschaftler (Gesamt 1,9%, H&F 4,7%) doppelt so häufig von Nutzern im Bereich Hochschule und Forschung aufgerufen als von der Nutzern in der Gesamtstichprobe. Umgekehrt fanden Seiten für Auszubildende bei den „akademischen“ Nutzern weniger Resonanz.

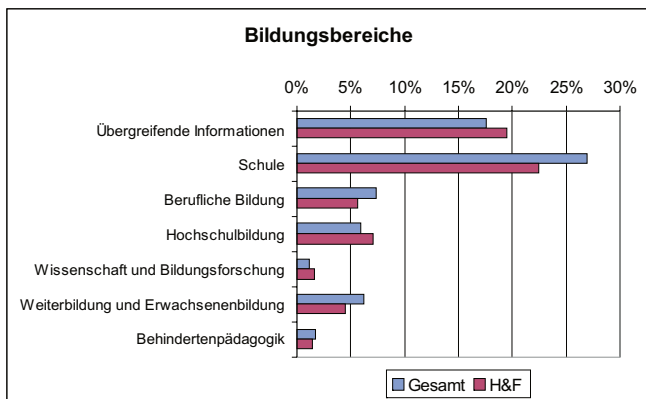


Abbildung 6

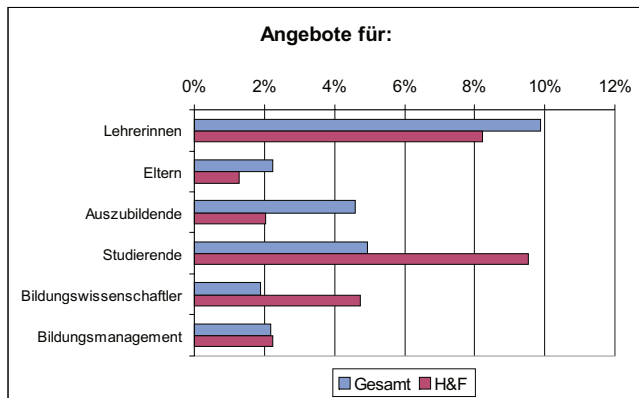


Abbildung 7

Zusammenfassung und Diskussion

In einer ersten nicht-repräsentativen Untersuchung hatten Vertreter unterschiedlicher Nutzergruppen bei Online Befragungen und Usability Untersuchungen der Qualität der Angebote des Deutschen Bildungsservers ein in wesentlichen Aspekten positives Votum erteilt (Brenstein & Kos, 2004). In dieser Untersuchung wurde der Versuch unternommen, das *tatsächliche* Nutzerverhalten anhand von Logfile-Daten eines Jahres (Okt. 01 - Sept. 02) zielgruppenspezifisch zu untersuchen. Da sich für die Zielgruppe der Nutzer aus Hochschule und Forschung eine Identifizierung über Hostadressen als Operationalisierungsheuristik anbietet, wurden die Zugriffe aller Nutzer mit Hostadressen aus dem Bereich Hochschule und Forschung selektiert und anhand einer vergleichenden Logfile-Analyse mit den Zugriffen der restlichen Nutzer verglichen. Dabei war von Interesse a) welchen Anteil diese so operationalisierte Stichprobe ausmacht, b) welche Unterschiede sich bei allgemeinen Nutzungsparametern ergeben und c) ob es differenziertere Zugriffe auf einzelne Themenbereiche gibt.

Die Ergebnisse zeigen, dass die Nutzung durch die Hochschul- und Forschungsstichprobe nur 6% der Gesamtnutzung ausmacht. Jedoch werden in dieser Gruppe verhältnismäßig mehr *unterschiedliche* Seiten genutzt. Dies lässt auf eine inhaltlich differenziertere Nutzung der Angebote des Deutschen Bildungsservers durch Besucher aus dem Bereich Hochschule und Forschung schließen. Bei den allgemeinen technischen Nutzungsparametern gab es dagegen kaum Unterschiede. In beiden Gruppen war Windows das am häufigsten genutzte Betriebssystem, jedoch war der allgemeine Browser-Favorit Internet-Explorer bei Besuchern aus dem Wissenschaftsbereich deutlich weniger beliebt.

Weiterhin zeigen die Ergebnisse, dass die Zugehörigkeit zu der einen oder anderen Stichprobe keinen Einfluss darauf hatte, wie viele Bildungsserverseiten aufgerufen wurden, da der Anteil der internen Referrer in beiden Stichproben etwa gleich war. Der hohe Anteil von Suchmaschinen bei den externen Referrern in beiden Gruppen zeigt, dass die Seiten des Bildungsservers am häufigsten über Suchmaschinen aufgerufen werden (bei Leuten, die den Bildungsserver nicht über Lesezeichen oder Adresseneingabe besuchen).

Das inhaltliche Interesse der Suchanfragen galt im Untersuchungszeitraum in beiden Gruppen verstärkt den Ergebnissen der Pisastudie. Unter den 30 am häufigsten geäußerten Suchanfragen waren außerdem Suchworte zum Deutschen Bildungsserver selbst sowie Anfragen zu Jobs und allgemeinen Themen in den Bereichen Schule und Hochschule. Dabei wurden hochschulrelevante Inhalte erwartungsgemäß deutlich häufiger von den Nutzern der Hochschul- und Forschungsstichprobe angefragt.

Bei der Nutzung der nach Bildungsbereichen gegliederten, redaktionell gepflegten Themenbereiche waren kaum stichprobenspezifische Unterschiede zu verzeichnen. Am stärksten nachgefragt, waren in beiden Gruppen die Bereiche

„Übergreifende Informationen“ und „Schule“. Dies ist unter anderem darauf zurückzuführen, dass diese Bereiche eine Fülle von Informationen anbieten und mehr Inhalte vorhalten als andere Bereiche. Zum anderen werden einige dieser Inhalte auch sehr häufig aufgerufen, da sie sowohl für akademische Nutzer als auch Vertreter anderer Bereiche und Tätigkeitsfelder relevant sind. Die Nutzungshäufigkeit einzelner Seiten wird in einem ausführlichen Projektbericht detailliert werden.

Trotzdem in den Bildungsbereichen keine stichprobenspezifischen Nutzungshäufigkeiten zu erkennen waren, weisen die deutlich unterschiedlich hohen Zugriffszahlen auf die zielgruppenspezifischen Angebote im Bereich „Angebote Für“ darauf hin, dass die Operationalisierung der Hochschul- und Forschungsstichprobe gelungen ist. Es zeigt sich, dass die Informationen für Studierende und Bildungswissenschaftler von Nutzern in der Hochschul- und Forschungsstichprobe stark frequentiert wurden. Diese Ergebnisse sind jedoch nur bedingt trennscharf und aussagekräftig, da Mitglieder dieser Zielgruppen auch über andere Service Provider auf die Seiten des Deutschen Bildungsservers zugegriffen haben können.

Auch bleiben nach der Betrachtung von Nutzerstatistiken noch viele Fragen offen. Zugriffszahlen liefern zwar einen Anhaltspunkt dafür, dass Angebote nützlich sind, vermitteln jedoch keinen Einblick in die vielfältigen Beweggründe und Einschätzung von Nutzern und Nicht-Nutzern von Online Bildungsinformationen. Daher kann nur eine repräsentative Befragung der relevanten Zielgruppen beantworten, ob und inwieweit, das, was Nutzer vorfinden, ihren spezifischen Ansprüchen genügt, welche Leerstellen aus ihrer Sicht bestehen und wie ein Bildungsportal in allen seinen Facetten rezipiert wird. Die derzeit laufenden Nutzerbefragungen und detaillierten Untersuchungen des Navigationsverhaltens sollen darüber weiterführend Auskunft geben und einen Beitrag zur marktgerechten Weiterentwicklung und Qualitätssicherung der Angebote des Deutschen Bildungsservers leisten.

Literatur

- Brenstein, E. Kos. O. (2004). Evaluation von Bildungsportalen - Empirische Untersuchungen zur Nutzung des Deutschen Bildungsservers. In: B. Bachmair/ P. Diepold/ C. de Witt (Hrsg.). Jahrbuch Medienpädagogik 4. Leverkusen: Leske+Budrich Verlag.
- Krug, Steve (2002). Don't make me think!: Web-Usability - Das intuitive Web. Landsberg: mitp.
- Nielsen, J. (2001). Designing Web Usability. München, Markt+Technik.

vascoda

Das gemeinsame Portal von Informationsverbünden, Elektronischer Zeitschriftenbibliothek und Virtuellen Fachbibliotheken

Christine Burbliès, Tamara Pianos

vascoda Geschäftsstelle c/o: Universitätsbibliothek Hannover und
Technische Informationsbibliothek (UB/TIB)

Abstract

Vascoda is a new portal for scientific information that is online since August 2003. (www.vascoda.de) It is a central access point for all fields of research. The portal allows interdisciplinary searches as well as (in a later version) complex navigation and browsing. Vascoda allows access to all types of documents: born-digital as well as digitised and print materials. The service includes full-texts, link-collections, databases, subject-specific search engines and more. The portal is also a nucleus in the creation of a German Digital Library.

The BMBF (Federal Ministry for Education and Research) and the DFG (German Research Society) are working together to make this portal a reality and to benefit from a joint venture instead of sponsoring two or more competing portals.

More than thirty German institutions – with co-operation partners worldwide – are working together to realize an actual one-stop-shop for all scientific information. The partners of this project are libraries as well as information centres and other institutions providing quality academic information.

The first release was realized through the co-operation of the network of subject-based Virtual Libraries, the Electronic Journals Library as well as the Information Alliance. In a joint effort, these institutions make it possible for users to combine the search for information with the access to full-text documents.

The central idea is to make it easy to obtain trustworthy quality information through a variety of complex searching and browsing options.

This paper will introduce the aims and services of vascoda and of the elements that contribute to vascoda.

1 Einleitung

Das Bundesministerium für Bildung und Forschung (BMBF) und die Deutsche Forschungsgemeinschaft (DFG) fördern seit Ende der 90er Jahre mehrere Einzelprojekte zur Verbesserung der Fachinformation. Die DFG-geförderten Virtuellen Fachbibliotheken sollen alle relevanten Informationen ihres Faches – ganz gleich in welcher Form sie vorliegen – nachweisen und zugänglich machen. Der Schwerpunkt liegt bei den BMBF-geförderten Informationsverbünden in der Verknüpfung von Recherche und elektronischem Volltext.

Ferner liefert die von DFG und BMBF gemeinsam geförderte Elektronische Zeitschriftenbibliothek (EZB) relevante Lizenzinformationen für elektronische Zeitschriften.

Seit August 2003 sind die Elektronische Zeitschriftenbibliothek, Informationsverbünde und Virtuellen Fachbibliotheken mit einem gemeinsamen Portal online, um so den Zugriff auf Fachinformation entscheidend zu verbessern. Das gemeinsame Portal vascoda (www.vascoda.de) ermöglicht sowohl eine interdisziplinäre Suche als auch eine komfortable Navigation zu den Einzelangeboten.

1.1 Zur Geschichte

Im Jahr 2001 gab es eine Kooperationsvereinbarung zwischen DFG und BMBF, da große Überschneidungsbereiche der jeweiligen Fördergebiete offenkundig wurden. Man beschloss die Energien zu bündeln und zukünftig gemeinsam an einem Portal für Fachinformation zu arbeiten anstatt zwei oder mehr konkurrierende Produkte zu fördern.¹ Um Doppelarbeiten zu vermeiden und Synergien zu

1 Zum Hintergrund und den frühen Entwicklungen vgl. DFG: Weiterentwicklung der überregionalen Literaturversorgung - Memorandum von 1998: ZfBB 45 (1998) 2, S. 135-164. Oder online unter http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/memo.pdf; Reinhard Rutz: Positionen und Pläne der DFG zum Thema Virtuelle Fachbibliothek. In: ABI-Technik 18, 1998, Nr. 4, S. 402-409, ders. SSG-Programm, Virtuelle Fachbibliotheken und Förderkonzept der DFG, online verfügbar unter: <http://webdoc.gwdg.de/ebook/aw/ssgfiwork/rutz.htm>; Jürgen Bunzel: Building the Digital Library in Germany: An Overview. In: LIBER QUARTERLY 8 1998, S. 23-37; Sven Meyenburg: Virtuelle Fachbibliotheken - fachspezifische Portale unter einem gemeinsamen Dach. In: 91. Deutscher Bibliothekartag in Bielefeld 2001. „Bibliotheken - Portale zum globalen Wissen“. Frankfurt a.M. 2001 (ZfBB Sonderheft 81), S. 30 - 36 und ders. Der Aufbau Virtueller Fachbibliotheken in der Bundesrepublik Deutschland. In: Bibliotheksdienst 34 (2000), Heft 7/8, S. 1229 - 1235.; Wilfried Enderle: Quo vadis SSG? Die deutsche Sondersammelgebietsbibliothek auf dem Weg zur virtuellen Fachbibliothek, online verfügbar unter: <http://www.archive.nrw.de/dok/workshop01/enderle/referat4.html>; Ewald Brahm: Aufbau und Förderung Virtueller Fachbibliotheken durch die Deutsche Forschungsgemeinschaft. Spezialbibliotheken heute - Wettbewerb und Kooperation. 28. Arbeits- und Fortbildungstagung der ASpB. Hannover, 6.-9. März 2001. Jülich, ASpB-Geschäftsstelle, 2001. 281-87.

nutzen, beschlossen im August 2002 alle Beteiligten, ihre Angebote in einem gemeinsamen Portal zu bündeln.

Ein zentrales Anliegen der gemeinsamen Anstrengungen ist die Optimierung der Informationsversorgung für die Nutzerinnen und Nutzer. „Der Mensch mit seinen spezifischen Anforderungen an wissenschaftliche Information muss im Vordergrund der Förderpolitik stehen.“² Damit wird dem Wunsch der Förderer entsprechend ein Paradigmenwechsel von der Anbieter- zur Nutzerseite vollzogen.

1.2 Die Ziele

„Mit dem Aufbau eines interdisziplinären Internetportals für wissenschaftliche Information in Deutschland soll ein transparenter, übergreifender Zugang zu allen Datenbanken, Volltexten und Dienstleistungen der überregionalen deutschen Informationseinrichtungen geschaffen werden.“³ Dieses Portal – vascoda – soll Nutzerinnen und Nutzern eine Orientierung in der Fülle der Webangebote bieten. Der Einstieg gewährleistet – anders als herkömmliche Suchmaschinen und die meisten Webkataloge – den Zugriff ausschließlich auf qualitätsgeprüfte Information. Außerdem erlaubt vascoda durch die Suche in Datenbanken eine Suche in Bereichen, die von herkömmlichen Suchmaschinen wie Google nicht erschlossen werden. Vascoda bietet somit also Zugriff auf das sogenannte „hidden web“.

2 Die Partner

2.1 Die Elektronische Zeitschriftenbibliothek

Die Elektronische Zeitschriftenbibliothek (<http://rzblx1.uni-regensburg.de/ezeit/>) bietet einen schnellen, strukturierten und einheitlichen Zugriff auf wissenschaftliche Volltextzeitschriften. Sie umfasst im April 2003 bereits 14.104 Titel, davon 1.655 reine Online-Zeitschriften, zu allen Fachgebieten. 4.090 Fachzeitschriften sind im Volltext frei zugänglich. Die an der Elektronischen Zeitschriftenbibliothek beteiligten Einrichtungen bieten den Zugriff auf die Volltexte der jeweils von ihnen abonnierten E-Journals. In das gemeinsame Portal vascoda bringt die EZB u.a. ihre Lizenzverwaltung ein, so dass Nutzerinnen und Nutzer von vascoda im Anschluss an eine Recherche direkt auf den elektronischen Volltext zugreifen können bzw. erfahren, wo die gewünschte Zeitschrift lizenziert ist. Kann auf diesem Wege keine Ausgabe des gewünschten Doku-

2 Strategisches Positionspapier des BMBF „Information vernetzen – Wissen aktivieren“ online unter: http://www.bmbf.de/pub/information_vernetzen-wissen_aktivieren.pdf, S. 2.

3 http://www.bmbf.de/pub/information_vernetzen-wissen_aktivieren.pdf, S. 10.

menten gefunden werden, wird die Nutzerin oder der Nutzer an einen Informationsverbund mit einer Pay-per-View-Option weitergeleitet.

2.2 Informationsverbünde

Ein Informationsverbund ist eine Kooperation zwischen Datenbankanbietern, Bibliotheken, Forschungseinrichtungen etc. eines Fachgebiets oder mehrerer Fachgebiete, die gemeinsam eine Dienstleistung zur Literatur- und Informationsversorgung für dieses Fach /diese Fächer aufbauen und betreiben. Dabei sind der Aufbau und das Angebot von elektronischen Volltexten und die Verknüpfung von Recherche und Volltext von zentraler Bedeutung. Die vier vom BMBF geförderten Informationsverbünde sind:

EconDoc (Informationsverbund Wirtschaft) www.econdoc.de wird realisiert durch das GBI in München, das Hamburgische Welt-Wirtschafts-Archiv (HWWA) in Hamburg, die Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW) in Kiel und die Universitäts- und Stadtbibliothek (USB) Köln. Die führenden Einrichtungen wirtschaftswissenschaftlicher Dokumentation haben im Informationsverbund EconDoc mehr als zwei Millionen Zeitschriftenartikel als Referenzen für Wirtschaftswissenschaft- und Praxis erschlossen.

GetInfo (Informationsverbund Naturwissenschaften und Technik) www.getinfo-doc.de wird realisiert durch das Fachinformationszentrum (FIZ Karlsruhe) und die Technische Informationsbibliothek (TIB) Hannover. GetInfo ermöglicht die Verknüpfung von Recherche und direktem Zugriff auf den elektronischen Volltext. GetInfo greift bei der Suche auf die TIBORDER-Datenbank sowie auf Verlagsdaten zurück und kann so u.a. auch graue Literatur, Forschungsberichte etc. nachweisen. Integriert ist Fachliteratur aus allen Bereichen der Technik und ihrer Grundlagenwissenschaften, insbesondere Chemie, Informatik, Mathematik und Physik, aber auch aus angrenzenden Bereichen der Naturwissenschaften.

infoconnex (Informationsverbund Pädagogik, Sozialwissenschaften, Psychologie) www.infoconnex.de wird realisiert durch das Deutsche Institut für Internationale Pädagogische Forschung in Frankfurt (DIPF), das Informationszentrum Sozialwissenschaften in Bonn (IZ) und das Zentrum für Psychologische Information und Dokumentation in Trier (ZPID). infoconnex durchsucht die führenden Datenbanken (FIS Bildung, Psyndex und SOLIS) im deutschsprachigen Bereich für Pädagogik, Psychologie und die Sozialwissenschaften und jeweils angrenzende Gebiete. Die Datenbanken weisen Zeitschriftenaufsätze, Aufsätze aus Sammelbänden, Monographien und graue Literatur nach.

Der Informationsverbund Medizin www.medpilot.de wird realisiert durch das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI) und die Deutsche Zentralbibliothek für Medizin (ZBMed) in Köln. MedPilot verknüpft die Recherche mit der Volltextbestellung bzw. in Kürze

auch mit Pay-per-View. In die Suche integriert werden können folgende Datenbanken und Kataloge: Medline, Medline Alert, Cochrane-Reviews, CCMed, Cancerlit, XToxline, Springer-Verlagsdatenbank, Kluwer-Verlagsdatenbank, Thieme-Verlagsdatenbank, Gerolit, Euroethics, Katalog: ZBMed Medizin, Ernährung/Umwelt, Katalog: Deutsche Zahnärztebibliothek, Katalog: NLM, Link-Datenbank der ZBMed, BMG-Pressemitteilungen, BGI-Presse Dienste, AnimAlt-ZEBET, EZB Regensburg (Online-Zeitschriften), Lehmanns Online Bookshop, CCRIS (Toxikologie), CIVS.

2.3 Virtuelle Fachbibliotheken

Die DFG möchte mit ihrem Förderprogramm „Elektronische Publikationen im Literatur- und Informationsangebot wissenschaftlicher Bibliotheken“ die Informationsversorgung in Deutschland verbessern.⁴ Einen besonderen Schwerpunkt des Förderprogramms bildet der Aufbau Virtueller Fachbibliotheken. Das organisatorische und fachliche Zentrum der Virtuellen Fachbibliotheken liegt fast immer bei den Sondersammelgebietsbibliotheken und Zentralen Fachbibliotheken und entspricht damit dem bislang geltenden Förderkonzept der DFG für gedruckte Materialien. Seit 1998 fördert die DFG einzelne Projekte zur Einrichtung Virtueller Fachbibliotheken; inzwischen gibt es bereits über 20 Projekte, an denen über 30 Einrichtungen beteiligt sind.

Anfang 2003 gibt es Virtuelle Fachbibliotheken zu folgenden Themen:

Online sind:

- Anglo-Amerikanischer Kulturraum Geschichte und Literatur (SUB Göttingen) *VLib Anglo-American Culture*
- Gegenwartskunst (SLUB Dresden) *ViFaArt*
- Geschichte (BSB München, SUB Göttingen, HU und StaBi Berlin), [in der Geschichte gibt es mehrere einzelne Virtuelle Fachbibliotheken: *Server Frühe Neuzeit*, *CLIO-Online* sind bereits online]
- Medizin (ZBMed Köln) *MedPilot*
- Niederländischer Kulturraum (ULB Münster) *NedGuide*
- Pharmazie (UB Braunschweig) *ViFaPharm*
- Physik (UB/TIB Hannover) *ViFaPhys*
- Politikwissenschaft und Friedensforschung (SUB Hamburg) *pp-guide*
- Psychologie (SULB Saarbrücken) *ViFaPsych*
- Sozialwissenschaften (IZ Bonn) *VIBSoz*
- Technik (UB/TIB Hannover) *ViFaTec*

4 http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/informationen_antragsteller/verteilte_digitale_forschungsbibliothek/elektron_publikationen.html

- Veterinärmedizin (BTiHo Hannover) *ViFaVet*
- Vorderer Orient (ULB Halle) *MENALIB*
- Wirtschaftswissenschaften BWL und VWL (USB Köln und ZBW Kiel) *EconBiz*

In Vorbereitung sind:

- Geowissenschaften, Geographie, Thematische Karten und Bergbau (SUB Göttingen)
- Geschichte Osteuropas (BSB München) *ViFaOst*
- Holztechnik (Dresden) *ViFaHolz*
- Ibero-Amerika (IAI Berlin)
- Ethnologie (HU Berlin) *EthnoViB*
- Rechtswissenschaften (SB Berlin) *ViFaRecht*
- Romanistik (Bonn)

Weitere Projekte befinden sich in der Antragstellungsphase.

Die geförderten Projekte waren zunächst hauptsächlich an den Interessen der jeweiligen Fachwissenschaftlerinnen und Fachwissenschaftler ausgerichtet und zeigen somit zwangsläufig ein sehr heterogenes Erscheinungsbild.⁵

Das Angebotsspektrum der Virtuellen Fachbibliotheken umfasst den Nachweis, die Bereithaltung, Erschließung und Archivierung von fachbezogenen elektronischen Informationsressourcen. Zum Angebot gehört aber auch, z.B. über die Integration relevanter OPACs, der Nachweis gedruckter Materialien. Diese können (in Zukunft) über einen integrierten Dokumentlieferdienst oder über Online-Fernleihe im Anschluss an die Recherche und ohne Wechsel des Anbieters bestellt werden. Des weiteren kann über ein durch einen Informationsverbund eingebundenes Pay-per-View-Verfahren der sofortige Zugriff auf Verlagstexte ermöglicht werden.

Virtuelle Fachbibliotheken überführen den umfassenden Sammelauftrag der SSG-Bibliotheken in das digitale Zeitalter und verknüpfen gedruckte und elektronisch vorliegende Materialien zu einer wahrlich hybriden Bibliothek. Dabei können durch den räumlich und zeitlich (außer durch Lizenzen) uneingeschränkten Zugriff erstmals Information zu dem jeweiligen fachlichen Inhalt des Sondersammelgebietes bereit gestellt werden, ohne dass ein Besuch der jeweiligen Bibliothek nötig ist.

Um Nutzerinnen und Nutzern die Vielfalt der nachgewiesenen Informationen und Dokumenttypen in einer übersichtlichen Form präsentieren zu können, sind die bislang existierenden Virtuellen Fachbibliotheken modular aufgebaut.

5 Viele der geförderten Projekte haben im Bibliotheksdienst ihr Angebotsspektrum dargestellt. Vgl. dazu die aktuelle Auflistung unter Punkt 4 im Newsletter No 7 der Virtuellen Fachbibliothek: http://www.virtuellefachbibliothek.de/newsletter_7.htm

In vielen Virtuellen Fachbibliotheken gibt es ähnliche Komponenten, die Zusammensetzung der einzelnen Module zu einem Gesamtprodukt variiert allerdings.

Die Angebote vieler Virtueller Fachbibliotheken enthalten Fachinformationsführer, fachspezifische Suchmaschinen, eine kombinierte Suche über Internetquellen und fachspezifische Datenbanken sowie Volltextserver z.B. für graue Literatur, aber auch Online Tutorials und Tagungskalender.

Fachinformationsführer / Internetquellennachweis / Guide

In den Fachinformationsführern sind Internetquellen und oft auch Fachliteratur nachgewiesen. Die Quellen sind intellektuell erschlossen und meist mit Kurzbeschreibungen versehen. In einigen Virtuellen Fachbibliotheken erfolgt die Erschließung der Quellen durch Fachwissenschaftlerinnen und Fachwissenschaftler, während in anderen die jeweils betroffenen Fachreferentinnen und Fachreferenten die Erschließungsarbeit leisten. Die Fachinformationsführer ermöglichen eine gezielte Navigation in genau begrenzten Gebieten. Hierbei kann meist zwischen einer fachlichen Untergliederung und einer Gliederung nach Ressourcentypen gewählt werden. Beim Fachinformationsführer Technik (FIT) können Filter gewählt werden, so dass man beispielsweise nur nach Tagungsankündigungen zum Thema Konstruktion suchen kann.⁶

Fachspezifische Suchmaschinen und Suchen über fachspezifische Angebote

Fachspezifische Suchmaschinen erleichtern die sachliche Suche im Internet, indem sie nur fachrelevante Sites indexieren. Beispiele für fachspezifische Suchmaschinen sind die Suchmaschine Technik und PsychSpider, die Suchmaschine des ZPID im Bereich Psychologie, die in das Angebot der Virtuellen Fachbibliothek Psychologie eingebunden ist.⁷

6 Beispiele für Fachinformationsführer: Fachinformationsführer Technik: <http://vifatec.tib.uni-hannover.de/fit/index.php3?L=g> ; Internetquellennachweis im NedGuide: <<http://suchfix.uni-muenster.de/vifanie/NedGuideNeu/index1.html>> ; Internetquellennachweis im pp-guide: <<http://www.pp-guide.de/intern/intern.shtml>>

7 Beispiele: Suchmaschine Technik: <<http://vifatec.tib.uni-hannover.de/sut/index.php3?L=g>> ; in die Virtuelle Fachbibliothek Psychologie ist die Suchmaschine des ZPID eingebunden: <<http://www.zpid.de/index.php?wahl=homep>> Eine Beschreibung der Funktionsweise der Suchmaschine Technik findet sich im Beitrag von Elzbieta Gabrys-Deutscher und Esther Tobschall „Kooperationen für die Virtuellen Fachbibliotheken“ im Bibliotheksdienst 36 (2002) H.8/9 unter http://bibliotheksdienst.zlb.de/2002/02_08_05.pdf

3 Ein gemeinsames Portal

Sowohl auf Seiten der Virtuellen Fachbibliotheken als auch auf Seiten der Informationsverbünde gab es bereits Planungen bzw. erste Schritte zur Umsetzung eines übergeordneten Portals. Diese Planungen werden in mehreren Schritten zusammengeführt, um von Anfang an mit einem einzigen übergeordneten Portal online zu gehen.

Die einzelnen Angebote sind bislang relativ heterogen. Einzelne Fächer wie Ingenieurwissenschaften, Anglistik, Medizin, Politikwissenschaften und Kunstgeschichte haben sehr unterschiedlich ausgeprägte Ansprüche an Informationen und ihre strukturierte Aufbereitung. Diesen vielfältigen Ansprüchen können Einzelangebote in vollem Umfange gerecht werden. Für ein übergeordnetes Portal jedoch, das eine übergreifende Suche und eine komfortable Navigation sowie ein fachliches Browsing ermöglichen soll, sind Anpassungen und Standardisierungen in einigen Bereichen unbedingt nötig. So müssen beispielsweise die Metadaten genau definiert und strukturiert werden. Dazu hat es auf der Ebene der Virtuellen Fachbibliotheken und der vom BMBF geförderten Informationsverbünde bereits jeweils Absprachen gegeben. Diese müssen nun wiederum angepasst werden, um das gemeinsame Portal Wirklichkeit werden zu lassen.

Eine Übersicht über das gesamte Angebot findet sich unter www.vascoda.de.

Die Integration weiterer Anbieter ist in einem zweiten Schritt vorgesehen. Vascoda ist prinzipiell offen für Erweiterungen.

Vascoda bietet einen interdisziplinären Einstiegspunkt, der sowohl eine übergreifende Suche als auch eine komfortable Navigation und (in einer späteren Version) ein Browsen im fachlichen Angebot erlaubt. Die Suche wird über Web-Services realisiert.⁸

Dieses Portal soll einen nutzergerechten Zugang bieten, der es ermöglicht, alle Informationsbedürfnisse an einem Ort zu befriedigen. Über vascoda soll die interdisziplinäre Suche in Datenbanken mit einem direkten Zugriff auf den gewünschten Volltext verbunden werden. Je nach physischer Form des Volltextes und der lizenzrechtlichen Situation kann der Volltext direkt aufgerufen, im Pay-per-View-Verfahren betrachtet oder über Online-Fernleihe oder andere Dokumentlieferdienste direkt bestellt werden.

In einem ersten Release, das zur IFLA 2003 online gegangen ist, ist zunächst eine interdisziplinäre Suche über das Angebot der Informationsverbünde sowie über viele Fachinformationsführer der Virtuellen Fachbibliotheken möglich.

8 Bei vascoda ist der Service Requestor das Portal vascoda. Die Informationsverbünde und die Fachinformationsführer der Virtuellen Fachbibliotheken stellen die Service Provider dar. Die Informationsverbünde und die Fachinformationsführer bieten einen Service an, der in der Lage ist, eine SOAP-Nachricht zu empfangen und diese auf Basis von SOAP zu beantworten.

Auf die Zeitschriftenangebote der EZB wird zunächst von gegebener Stelle aus verlinkt, um so auf lizenzierte Volltexte bzw. auf die Lizenzinformationen zu gelangen. Außerdem wird zunächst über Links der Zugang zu allen anderen Angeboten der Projektpartner ermöglicht.

Für das zweite Release sind erweiterte Suchfunktionalitäten vorgesehen. So wird es in einem zweiten Release beispielsweise eine Filtermöglichkeit nach Medienform geben, um die Suche und die Ergebnisanzeige stärker differenzieren zu können.⁹

Kontakt

Christine Burbliès und Dr. Tamara Pianos
vascoda Geschäftsstelle c/o: Universitätsbibliothek Hannover und
Technische Informationsbibliothek (UB/TIB)
Postanschrift: Postfach 6080, D-30600 Hannover
Hausanschrift: Welfengarten 1 B, D-30167 Hannover
Tel.: +49 (0)511 762-4193 / Fax: +49 (0)511 762-2924
christine.burbliès@tib.uni-hannover.de
tamara.pianos@tib.uni-hannover.de
<http://www.vascoda.de>

9 Zu den neueren Entwicklungen vgl. auch zwei Beiträge in ZfBB 50 (2003) Christa Schöning-Walter: Die Digitale Bibliothek als Leitidee: Entwicklungslinien in der Fachinformationspolitik in Deutschland, S. 4-12 und Uwe Rosemann: Die Arbeitsgruppe der Informationsverbünde und die Virtuellen Fachbibliotheken: Beginn einer wunderbaren Zusammenarbeit?!, S. 13-18.

Probleme der Integration digitaler Bibliothekssysteme: Semantische Heterogenität bei datenbankübergreifenden Recherchen

Klaus Hahn

Zentrum für Psychologische Information und Dokumentation
Universität Trier, D-54296 Trier, Deutschland

Zusammenfassung

Die Weiterentwicklung digitaler Bibliothekssysteme führt heute zu einer bisher nicht gekannten Integration wissenschaftlicher Recherche: Grenzen zwischen den Fachgebiets-spezifischen Literaturdatenbanken verschwimmen.

Ein gemeinsames Projekt der Fachinformationsanbieter dreier Disziplinen (Bildung: Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt/Main; Sozialwissenschaften: Informationszentrum Sozialwissenschaften, Bonn; Psychologie: Zentrum für Psychologische Information und Dokumentation, Trier) arbeitet seit einiger Zeit an dem Ziel, Benutzern über eine zentrale Recherche-Oberfläche einen integrierten Zugang zu allen gemeinsamen Daten - von bibliografischen Angaben bis hin zum online-Volltext - zu ermöglichen. Die Sondersammelgebietsbibliotheken der drei Fachgebiete übernehmen dabei die Lizenzierung von Zeitschriftenartikeln im Volltext. Das Projekt wird von BMBF und DFG gefördert.

Die Integration dieser drei Systeme hat zum Ziel, dem Nutzer die komfortable und effiziente Ausweitung seiner Recherche anzubieten. Dazu soll die qualitativ hochwertige Inhalterschließung der Fachdatenbanken genutzt werden, die auf der Basis fachspezifischer Thesauri und Klassifikationen vorliegt. Es müssen dabei aber Wege gefunden werden, semantische Mehrdeutigkeiten zu vermeiden - die mit Sicherheit auftreten würden, wenn Suchbegriffe eines Fachgebiets ohne adäquate Transformation zur Recherche in den anderen Fachgebieten verwendet würden. Diese Heterogenitätsproblematik stellt sich nicht nur zwischen Fachthesauri, sondern z.B. auch bei der Übersetzung von Fachbegriffen in andere Sprachen. Die Stabilität des semantischen Kontexts kann zum einen durch den Aufbau intellektueller Cross-Konkordanzen gewährleistet werden, zum anderen durch statistische Transferverfahren, die auf der Basis von Parallelkorpora unterschiedliche Vokabulare aufeinander abbilden.

Abstract

The advancement of digital library systems during the last years led to a so far unknown integration of scientific literature research: traditional borders between the established domain-specific literature databases blur today.

An integrated project of Germany-based information suppliers within three scientific disciplines (Education: German Institute for International Educational Research, Frankfurt/Main; Social Sciences: Information Center for the Social Sciences, Bonn; Psychology: Center for Psychological Information and Documentation, Trier) works for some time now towards offering to the scientific public one integrated front-end as an entrance to all of their data, ranging from simple bibliographic meta-data to complete on-line papers. So-called „Sonder-sammelgebietsbibliotheken“ of the three domains are responsible for acquiring and licensing of magazine articles and original papers. The project is funded by the „Bundesministerium für Bildung und Forschung (BMBF)“ and the „Deutsche Forschungs-Gemeinschaft (DFG)“.

An integration of these distributed databases tries to offer a comfortable and efficient expansion of a user's original, single-database search. To this end the already established knowledge structures within the domain-specific databases (so-called „Inhaltserschließung“) are to be used, which comprise thesauri and classifications. Methods are under development to avoid the semantic ambiguities which would surely arise would search terms of one domain be propagated to another one without adequate transformation. This so-called Problem of Semantic Heterogeneity is not only related to domain-specific thesauri but arises e.g. also during the translation of technical terms.

The goal is to try to ensure the stability of a given semantic context on the one hand by intellectually constructed cross-concordances built between two thesauri, on the other hand by correlation-based mappings, which use parallel corpora in order to obtain a concordance measure.

Problemstellung: Information Retrieval und Meta-Daten

Die traditionelle wissenschaftliche Recherche muss sich heute der Herausforderung durch die zwar chaotische, aber auch durchaus integrative Präsenz des Internet stellen. War es bislang üblich, wissenschaftliche Veröffentlichungen vornehmlich in darauf spezialisierten, verstreut aufgebauten Datenbanken mit jeweils eigener Zugangslogik zu suchen, so hat sich inzwischen - durch das starke Anwachsen wissenschaftlicher Publikationen einerseits und die explosionsartige Vermehrung so genannter grauer Literatur andererseits - das globale Durchstöbern des Internet als durchaus fruchtbar erwiesen, wobei sich die Ergebnisse

aber im Normalfall auf den kostenlos absuchbaren Teil des Internet beschränken.

Information Retrieval heißt in erster Linie *Text Retrieval*. Gesucht wird üblicherweise mit den jeweils effizientesten *Suchmaschinen*, im Augenblick z. B. häufig mit *Google*. Derzeit repräsentiert *Google* ungefähr 3 Milliarden Web-Pages und beantwortet täglich ca. 250 Millionen Suchanfragen. Über die Qualität der gefundenen Dokumente bzw. deren Passung zum intendierten Thema kann *Google* lediglich statistische Aussagen machen: Alle im Such-Corpus enthaltenen Dokumente werden nach bestimmten Heuristiken bzw. der Repräsentativität / Zentralität der enthaltenen Suchbegriffe gerangiert. Da die allermeisten Texte im Internet nicht semantisch qualifiziert sind - also keine *Meta-Daten* enthalten - haben Suchmaschinen auch keine andere Möglichkeit als die der Volltextsuche. Als Ergebnis einer Such-Anfrage erhält der Nutzer also eine eher heuristisch geordnete Menge von Dokumenten, die er nun entweder alle prüfen oder aber mit weiteren Spezifikationen von Suchbegriffen und entsprechenden Boole'schen Relationen einengen muss.

Dieses Fehlen von Meta-Daten im Internet wird inzwischen auch von den verantwortlichen Gremien als zentrales Problem erkannt (Semantic Web, [1]) und ist auf der Agenda immer weiter nach oben gewandert. Verschiedene Lösungsansätze werden seit einiger Zeit diskutiert (Dublin Core Metadata Initiative, [2]; Resource Description Framework, [3]; Ontology Layer, [4]), da aber parallel dazu eine weit reichende Diskussion über neue Strukturierungsmodelle (Extensible Markup Language, [5]) im Internet geführt wird, scheint zumindest für die nahe Zukunft keine Standardisierung in Sicht zu sein. Auch wird die Akzeptanz normierter Strukturen und damit deren Sinnhaftigkeit immer deutlicher generell infrage gestellt. Jede Kennzeichnung mithilfe von Meta-Daten setzt im Prinzip eine Übereinkunft über die zu verwendenden Terme voraus - und die Analyse der gegenwärtigen Ansätze im Hinblick auf Klassifikationen, Thesauri und Schlagwörter stimmt dabei alles andere als optimistisch [6]. Die historisch gewachsenen Regelwerke und Normen stammen aus einer Zeit, in der die im Wesentlichen wissenschaftlich tätige Klientel einerseits den großen Nutzen solcher Standards sah, als auch deren Verwaltung und Management durch kleine, elitäre Gremien akzeptierte.

Die Veränderung der Publikationslandschaft durch das Internet zeigt, dass derartige Strukturen überholt sind. Ein wortgewandter Kryptographie-Spezialist, der auf seiner Web-Site einen avancierten Aufsatz zur Datensicherheit im Internet veröffentlicht, wird wahrscheinlich nicht seine Zeit darauf verschwenden, die entsprechenden Verwaltungsgremien zu kontaktieren, um deren Ansicht zur Verschlagwortung seiner Arbeit einzuholen. Will man trotz solcher Entwicklungen nicht auf die Effizienz intellektuell erschlossener Datenbestände verzichten, ist ein mehrdimensionaler Ansatz der semantischen Qualifizierung unumgänglich („Standardisierung von der verbleibenden Heterogenität

her“ [6]). So wäre es durchaus fruchtbar, die verschlagworteten Archive durch automatische, heuristisch indexierende Verfahren zu ergänzen: Mit einiger Wahrscheinlichkeit werden in Zukunft automatische bzw. semi-automatische Verfahren schon deshalb eingesetzt werden müssen, um der Vielfalt und Wandlungsfähigkeit des Publikationsmarktes einigermaßen gewachsen zu sein.

Nun nimmt im Augenblick ein ‚normaler‘ Benutzer wahrscheinlich in Kauf, die von der Suchmaschine gelieferte, unqualifizierte Ergebnismenge quasi explorativ weiter einzugrenzen, da die aufgewendete Zeit meist nicht als kritisch empfunden wird. Der *professionelle* Benutzer (z.B. Wissenschaftler) aber, der mit wenig Aufwand maximale Treffergenauigkeit erreichen will und muss, kann mit dieser Unschärfe nicht ernsthaft arbeiten - zu diesem Zweck wurden ca. 1960 die so genannten *Fachdatenbanken* entwickelt. In ihnen sind wissenschaftliche Publikationen durch die wichtigsten Meta-Daten repräsentiert: *Quelle*, *Autor*, *Zusammenfassung*, *Schlagwörter*, etc. sind als solche gekennzeichnet und ermöglichen damit eine präzisierbare Suche. Diese Datensammlungen verwenden darüber hinaus (als sog. erschlossene Datenbasen) eine Kennzeichnung der Referenzen (*Verschlagwortung*) mithilfe eines fachspezifischen Vokabulars, des *Thesaurus*. *Thesauri* oder *Klassifikationen* sind dabei einerseits Basis für die Präzisierung semantischer Mehrdeutigkeiten: Der Benutzer kann davon ausgehen, dass der im Thesaurus definierte Begriff auch zur Indexierung aller zugehörigen, thematisch relevanten Veröffentlichungen herangezogen wurde. Außerdem dient die in einem Thesaurus definierte semantische Struktur eines Themengebiets zur schnellen Orientierung und damit auch zur schnellen Erweiterung oder Eingrenzung eines Suchbereichs.

Der wesentliche Nachteil dieses Aufbaus fällt schnell ins Auge: Es ist unumgänglich, dass der Nutzer diesen der Datenbasis zugrunde liegenden Thesaurus kennt, oder dass er bereit und in der Lage ist, ihn zu analysieren. Ist ein Thesaurus umfangreich (was meist der Fall ist: mehr als 5.000 Terme sind üblich, es können durchaus auch über 30.000 sein), wird die Analyse zeitaufwändig. Da aber Fachdatenbanken primär die Effektivität des Suchprozesses steigern sollten, kann einem Nutzer nicht ernsthaft zugemutet werden, die Thesauri mehrerer Fachvokabulare zu analysieren: Hier wäre selbst eine *Google*-basierte Variante effektiver.

Deshalb werden seit einiger Zeit koordinierte Anstrengungen unternommen, *semantische Übergänge* zwischen fachspezifischen Thesauri und zu allgemeineren Verzeichnissen, z.B. der Schlagwortnormdatei (SWD) der Deutschen Bibliothek, Frankfurt/Main, zu entwickeln. Gerade im Zeitalter des Internet, mit dessen Hilfe sich Benutzer einen *zentralen* Zugang zu verteilten Datenbeständen erhoffen, sollte sich niemand mehr mit der Komplexität isolierter Insel-Lösungen auseinander setzen müssen.

Projektansatz

Will man den professionellen Benutzer bedienen, dann sollte idealerweise die Durchsichtigkeit einer modernen, Internet-basierten Suchmaschine mit der Leistungsfähigkeit Metadaten-gestützter Fachdatenbanken verbunden werden. Ein derartiges System müsste einerseits in der Lage sein, eine Suchanfrage (in normierter Form) an alle gewünschten (Fach-) Datenbanken durchzureichen, andererseits sollte es dafür Sorge tragen, dass der vom Benutzer beabsichtigte semantische Kontext - wenn irgend möglich - in diesen Datenbanken so gut wie möglich rekonstruiert wird. Das Projekt des *Informationsverbundes Bildung – Sozialwissenschaften - Psychologie (BSP)* hat das Ziel, ein solches Portal zu entwickeln. Dabei wird

- zur System-neutralen Repräsentation der Anfrage eine einheitliche XML-Schnittstelle mit entsprechenden Protokollen entwickelt [7],
- die semantische Heterogenität durch den Aufbau von Konkordanz-Tabellen mit gewichteten Graphen („Cross-Konkordanzen“, [8, 9, 10]) reduziert,
- dem Nutzer die Möglichkeit gegeben, die gewünschten Publikationen aus der Suchmenge auf vielfältige Weise (online-reader, Ausdruck, kostenpflichtige oder kostenlose Anlieferung, etc.) zu erhalten.

Der Verbund setzt sich zusammen aus wissenschaftlichen Einrichtungen, die die thematischen Kompetenzen absichern, und Bibliotheken, die sich in erster Linie um das Handling der Publikationen und die rechtlichen Fragen kümmern:

- *Deutsches Institut für Internationale Pädagogische Forschung*, Frankfurt/Main (DIPF), mit einem Datenbestand von ca. 500.000 Literaturnachweisen ab 1980 und einem Thesaurus, der im Augenblick fast 60.000 Begriffe umfasst.
- *Informationszentrum Sozialwissenschaften*, Bonn (IZ), mit ca. 250.000 Literaturnachweisen ab 1945 und einem Thesaurus mit ca. 7.400 Termen.
- *Zentrum für Psychologische Information und Dokumentation*, Trier (ZPID), mit ca. 170.000 Literaturnachweisen ab 1977 und einem Thesaurus mit ca. 5.200 Termen.
- *Universitätsbibliothek Erlangen-Nürnberg* (UB Erlangen)
- *Universitäts- und Stadtbibliothek Köln* (USB Köln)
- *Saarländische Universitäts- und Landesbibliothek* (SULB)

Semantische Heterogenität: Bedeutung als Problem

Wird eine Recherche gleichzeitig in mehreren Datenbanken durchgeführt, so ist es vielleicht nicht unwahrscheinlich, dass der gesuchte Term in all diesen Systeme-

men auch als Begriff isoliert worden ist – er repräsentiert aber mit hoher Wahrscheinlichkeit nicht überall den gleichen *semantischen Raum* (z.B. „Statistik“, womit die Psychologie in erster Linie *Wahrscheinlichkeitstheorie* meint, die Soziologie aber eher *Demoskopie*). Mit Sicherheit wird es nicht oft eine *direkte* semantische Umsetzung von einem Wissensbereich in einen anderen geben – jedes Fachgebiet hat eigene, historisch gewachsene Strukturen, mit deren Hilfe es das repräsentierte Wissen abzubilden versucht. Probleme der Granularität (z.B. der Begriffe *Psychotherapie* oder *Angst* im IZ- vs. ZPID-Thesaurus) mischen sich mit Schwierigkeiten wie Synonym-Verwendung (z.B. *Aufgeschlossenheit* vs. *Offenheit*) und Homonymen (*Verband*, *Bank*, etc.). Darüber hinaus unterscheiden sich Thesauri auch im Hinblick auf die sog. *Koordiniertheit*: *Prä-kordinierte* Thesauri verwenden zusammengesetzte Terme (*Jugendarbeitslosigkeit*), während *post-kordinierte* Thesauri die Basis-Terme trennen (*Arbeitslosigkeit*, ..., *Jugend*) und deren Zusammenführung in einer *und-ierten* Suche erwarten.

Die Analyse der im Wesentlichen beteiligten Thesauri (DIPF, IZ, SWD [= Schlagwort-Normdatei der Deutschen Bibliothek, Frankfurt/Main], ZPID) zeigt, mit welchen Differenzierungen und Eigenarten hierbei umgegangen werden muss:

- Bereits bei den Schreibweisen gibt es verschiedene Ansätze, die noch nicht einmal nur der *Neuen Deutschen Rechtschreibung* geschuldet sind, sondern durchaus eigene Interpretationen darstellen, beispielsweise:
 - ZPID: Clusteranalyse - SWD: Cluster-Analyse
 - ZPID: Creutzfeldt-Jakob-Syndrom - SWD: Jakob-Creutzfeldt-Syndrom
 - ZPID: Cykloheximid - SWD: Cycloheximi
 - ZPID: Tranquilizer - DIPF: Tranquillizer - SWD: Tranquillizer
 - ZPID: Känguruhs - SWD: Känguru
- Die größte Gruppe stellen Begriffe dar, die im Ziel-Thesaurus zwar eine gute bis sehr gute semantische Entsprechung haben, welche aber nicht durch einen einfachen algorithmischen Prozess gefunden werden kann:
 - ZPID: Abtreibungsgesetze - IZ: Schwangerschaftsabbruch
 - ZPID: Lob - IZ: soziale Anerkennung
 - ZPID: Lupus - IZ: Hautkrankheit
 - ZPID: Masochismus - IZ: Sexualstörung
 - ZPID: Virginität - DIPF: Jungfraulichkeit
 - ZPID: Vigilanz - DIPF: Aufmerksamkeit
- Ein weiteres Cluster beinhaltet Terme, die in einer Art Ober-/Unterbegriff zueinander stehen - hier gilt es oft, weitere semantische Dimensionen zu berücksichtigen, was eine gute Kenntnis beider Wissensbereiche notwendig macht:
 - ZPID: Marihuanakonsum - IZ: Drogenmissbrauch

- ZPID: Nachbild - IZ: visuelle Wahrnehmung
- ZPID: Recency-Effekt - SWD: Positionseffekt
- ZPID: Wernicke-Syndrom - DIPF: Sprachstörung
- Damit in engem Zusammenhang finden sich Begriffe, deren Semantik zwar eine Art Ordnungsrelation nahe legt, die aber nicht durch eine einfache Über-/Unterordnung repräsentierbar ist:
 - ZPID: Kastrationsangst - IZ: Psychoanalyse
 - ZPID: Introjektion - IZ: Psychoanalyse
 - ZPID: Katharsis - IZ: Psychoanalyse

Unter solchen Voraussetzungen ist es keine triviale Aufgabe, den vom Benutzer intendierten semantischen Raum adäquat zu transferieren. Dieser Transfer ist nun das Anliegen der Behandlung semantischer Heterogenität: Die Such-Anfrage eines Benutzers muss – ohne dessen Zuarbeit - so effizient wie möglich gegen viele verschiedene (Fach-) Datenbanken gestellt werden können.

Lösungsmodelle

Die bisherigen Ansätze, um eine semantische Umsetzung vom Ausgangs- in den Zielraum zu konstruieren, sind

- Statistische Modelle (Häufigkeitsanalysen, Co-Occurrence Analysis / Bedingte Wahrscheinlichkeiten, [11]).
- Neuronale Netzwerk-Modelle (prinzipiell analog den Inferenzstatistischen Modellen: Multiple Regression z.B. über Hidden Layer, [11]).
- Intellektuell erarbeitete Konkordanzen (gewichtete Graphen als Übergang zwischen Thesauri: „Cross-Konkordanzen“, [7]).

Klassische Inferenzstatistische Modelle, wie auch ihre modernen, neuronal modellierenden Nachfolger, arbeiten prinzipiell mithilfe der Analyse bedingter Wahrscheinlichkeiten: Über das gemeinsame Auftreten von Termen aus verschiedenen Thesauri bei der Verschlagwortung einer bestimmten Publikation schließen sie auf semantische Zusammenhänge zwischen eben diesen Termen. Als Erweiterung der gebräuchlichen statistischen Ansätze (z.B. der Multiplen Regressionsanalyse), die das Datenmaterial in geschlossener Form analysieren, modellieren die sog. *Neuronalen Netzwerke* mithilfe inkrementellen *Lernens*: idealerweise entsteht dabei eine robuste und flexible Repräsentation der Zusammenhänge in den Daten.

Maschinenbasierte Ansätze haben natürlich den großen Vorteil automatischer Durchführbarkeit und scheinen im Hinblick auf Wartungs- und damit Kosteneffizienz die bessere Wahl zu sein. Um aber zufrieden stellende Ergebnisse liefern zu können, sind jedoch große Mengen *parallel verschlagworteter*

Arbeiten notwendig (> 50.000), ansonsten stören zufällige Zusammenhänge im Datenmaterial (z.B. zeitgeschichtliche Besonderheiten) die Repräsentativität der Ergebnisse ganz erheblich. Die dann ermittelten „semantischen Korrelationen“ sind kaum verwendbar, sie reflektieren lediglich Trends in der aktuellen Publikationslage. Automatische Verfahren liefern auch keine Informationen über die *Art* der semantischen Relation: Ob es sich um ein Ober-/Unterbegriffs-Paar oder etwa um äquivalente Formen handelt, müsste ein menschlicher Bearbeiter hinzufügen. Auch um sicher zu stellen, dass die statistisch ermittelten Konkordanzan inhaltlich nicht allzu divergent sind, wären Kontrollen durch Fachleute wünschenswert.

Die *intellektuelle* Zuordnung von Termen auf der anderen Seite ist arbeits- und kostenaufwendig, und das nicht nur bei der Erstellung, sondern als permanenter Faktor: Da sich Wissenschaftsthemen und damit Thesauri ständig ändern, muss eine Konkordanz fortlaufend durch qualifizierte Fachkräfte aktualisiert bzw. gepflegt werden. Dafür stellt sie so qualitativ das Optimum der Möglichkeiten dar: Fachleute aus den beteiligten Wissensgebieten einigen sich auf die semantischen Räume, zwischen denen sinnvolle Transfers über Term-Zuordnungen konstruiert werden.

Wegen dieser qualitativen Vorteile entschied man sich auch beim Informationsverbund BSP für die Implementierung intellektueller Konkordanzen. Zur konkreten Realisation wurde auf die Vorarbeiten des CARMEN-Projekts [8,9] zurückgegriffen, in dessen Umfeld mehrere Ansätze untersucht und getestet worden waren. Übernommen wurde dabei der CARMEN-Ansatz zur Implementation von Konkordanzen als *gewichtete Graphen*, die zwei dreiwertige Informationen repräsentieren:

1. die *Art* der Relation
 - a. Äquivalente Terme { = }
 - b. Verwandte Terme { ^ }
 - c. Ober- bzw. Unterbegriff { >; < }
2. die *Stärke* (das *Gewicht*) dieser Relation
 - a. gering
 - b. mittel
 - c. hoch

Die einfachste Form einer Konkordanz lässt sich dabei als 1:1 - Relation darstellen, die qualitativ noch zwischen *Äquivalenz* und *Verwandschaft* unterscheidet, wobei es vielleicht etwas seltsam erscheint, eine *Äquivalenz* zu gewichten:

Sozialpsychiatrie {=;h} → Sozialpsychiatrie

Übung {^;h} → Training

Zusätzlich können und müssen die Zuordnungen fallweise mithilfe von und/oder - Verknüpfungen über die einfache 1:1 - Relation hinaus erweitert werden:

Arbeitnehmermotivation {^+;m} → **Arbeitnehmer** + **Motivation**

(d.h. *Arbeitnehmermotivation* aus Thesaurus A ist semantisch verwandt mit der und-ierten Verknüpfung der beiden Terme aus Thesaurus B, *mittlere Relevanz*).

Motivation {>o;h} → **Intrinsische Motivation**

{>o;h} → **Extrinsische Motivation**

{>o;h} → **Leistungsmotivation**

(*Motivation* ist der Oberbegriff für die oder-ierten drei Terme aus Thesaurus B, jeweils *hoch relevant*).

Introjektion {^;m} → **Psychoanalyse**

Kastrationsangst {^;h} → **Psychoanalyse**

Katharsis {^;h} → **Psychoanalyse**

(*Psychoanalyse* aus Thesaurus B ist semantisch verwandt mit einer großen Anzahl von Begriffen aus dem offensichtlich in dieser Hinsicht präziser auflösenden Thesaurus A).

Sucht ein Benutzer nun z.B. nach Arbeiten zum Thema *Motivation*, dann kann er im projektierten System diese Suche z.B. auf eine weitere Ziel-Datenbank ausdehnen und erhält u. a. die (*Meta-*) Information, dass der Suchbegriff im Ziel-Thesaurus den *Oberbegriff* für *drei spezialisiertere Terme* darstellt, die als für die Suche *hoch-relevant* erachtet werden. Er kann sich dann entscheiden, ob er mit dieser *oder-ierten* Zielmenge weiterarbeiten möchte, oder die Suche auf eine Teilmenge dieser Begriffe eingrenzen.

Im Endeffekt hat der Nutzer also ein intelligent erweitertes Angebot - es geht nicht darum, ihn mit Ergebnissen zu konfrontieren, die seine Such-Intention überinterpretieren. Durch die als eine Art Zwischenschicht zu betrachtenden Konkordanz-Tabellen kann eine beliebig komplexe Term-Transformation vorbereitet werden, es liegt dann am Nutzer, darauf aufbauend weiterzuarbeiten. Ob und mit welchem Erfolg solche semantischen Pfade akzeptiert werden, wird natürlich entscheidend für die Zukunft des Ansatzes sein. Idealerweise könnte dem System eine Art „Rückkopplung“ hinzugefügt werden, mit deren Hilfe die Nutzer an der Qualität der Term-Propagierung mitwirken könnten - im Internet arbeiten inzwischen einige Portale mit „Nutzer-Empfehlungen“, und die Akzeptanz scheint sehr gut zu sein.

Fazit

Zur effektiven Lösung des Transfer-Problems bei semantisch heterogenen Thesauri wurde im Verbundprojekt *Bildung-Sozialwissenschaften-Psychologie* die Konstruktion *intellektuell optimierter Konkordanzen* gewählt. Die Konkordanz-Tabellen sind inzwischen fertig gestellt und die während der Erarbeitung stichprobenhaft durchgeführten Validierungen zeigen klare Vorteile des Ansatzes: Die in Ziel-Thesauri über Konkordanzen aufgefundenen Publikationen repräsentieren natürlich deutlich präziser die ursprüngliche Such-Anfrage, als es eine einfache Term-Weitergabe hätte leisten können.

Mit hoher Wahrscheinlichkeit stellt der bereits erwähnte Aufwand bei der Konstruktion und Wartung von Konkordanzen zumindest in Zukunft ein ernstes Problem dar, weil prinzipiell zwischen allen relevanten Thesauri - und sinnvollerweise auch zwischen verschiedenen Sprachen - semantisch vermittelt werden kann. Da sich Thesauri darüber hinaus permanent den wissenschaftlichen Strukturveränderungen anpassen müssen (oft werden pro Jahr mehr als 10% der Terme umstrukturiert, ausgetauscht, gelöscht, ergänzt), sollten intellektuelle Konkordanzen genauso regelmäßig überarbeitet werden. Dies könnte aber nur durch qualifizierte Fachkräfte übernommen werden, die über den Themenbereich eines einzigen Fachgebiets hinaus sehen müssen - solche Fachleute sind nicht billig.

Welche Methoden sich in der vernetzten Wissenschaftswelt der Zukunft zur Heterogenitätsbehandlung durchsetzen können, ist offen - vielleicht wird die aktuelle Forschung zur automatischen Indexierung schriftlicher Arbeiten hier ein wertvoller nächster Schritt. Obwohl die maschinenbasierte semantische Analyse nicht-trivialer Texte bisher keine wirklich brauchbaren Ergebnisse hervorgebracht hat, dürften sich solche Ansätze aus den erwähnten Kostengründen mehr und mehr in den Vordergrund drängen. Einverständnis herrscht im Augenblick offensichtlich nur darüber, dass die traditionelle, administrative Normung der wissenschaftlichen Publikationswelt den kommenden Herausforderungen nicht mehr gewachsen ist [6]. Weltweit wird deshalb angestrengt darüber nachgedacht, wie der semantische Gehalt des größten Wissensspeichers aller Zeiten - des Internet - nutzbar gemacht werden kann.

Welche Wege auch immer beschritten werden - im Endeffekt entscheidet die *Qualität* der Lösung, also der *Nutzer-Mehrwert*, über das Schicksal integrierter Such-Ansätze: Gelingt es nicht, dem *Kunden* einen spürbaren Vorteil zu liefern, dann bleiben alle Entwicklungen auf bloße akademische Wissenserweiterung begrenzt.

Literatur

- [1] Semantic Web (www.w3.org/2001/sw/)
- [2] Dublin Core Metadata Initiative (dublincore.org/)
- [3] Resource Description Framework (www.w3.org/RDF/)
- [4] WebOnt (www.w3.org/2001/sw/WebOnt/)
- [5] XML (www.w3.org/XML/)
- [6] Krause, J., Niggemann, E., Schwänzl, R. (2003): Normierung und Standardisierung in sich verändernden Kontexten: Beispiel Virtuelle Fachbibliotheken. In: ZfBB: Zeitschrift für Bibliothekswesen und Bibliographie 50, Nr. 1, S. 19 - 28.
- [7] Strötgen, R. (2002): Behandlung semantischer Heterogenität durch Metadatenextraktion und Anfragetransfer. In: Womser-Hacker, C., Wolff, C., Hammwöhner, R. (Hrsg.): Information und Mobilität, Proceedings des 8. Internationalen Symposiums für Informationswissenschaft (ISI 2002). Konstanz: UVK (Schriften zur Informationswissenschaft, Band 40).
- [8] www.gesis.org/Forschung/Informationstechnologie/CARMEN-AP12.htm
- [9] www.gesis.org/Forschung/Informationstechnologie/CARMEN-AP11.htm
- [10] Binder, G., Marx, J., Mutschke, P., Riege, U., Strötgen, R. (Informationszentrum Sozialwissenschaften); Kokkelink, S., Plümer, J. (Universität Osnabrück) (2002): Heterogenitätsbehandlung bei textueller Information verschiedener Datentypen und Inhaltserschließungsverfahren. (Arbeitsbericht Nr. 24, IZ Sozialwissenschaften, Bonn).
- [11] Hellweg, H., Krause, J., Mandl, Th., Marx, J., Müller, M.N.O., Mutschke, P., Strötgen, R. (2001): Treatment of Semantic Heterogeneity in Information Retrieval. (Arbeitsbericht Nr. 23, IZ Sozialwissenschaften, Bonn).

Unterstützung kooperativer Verfahren zum Aufbau von Fachportalen

Heiko Hellweg, Bernd Hermes und Maximilian Stempfhuber

Informationszentrum Sozialwissenschaften (IZ), Lennéstr. 30, 5313 Bonn
stempfhuber@iz-soz.de

Abstract

Fachinformationsführer, Clearinghouses oder Subject Gateways, wie sie zum Beispiel in den Virtuellen Fachbibliotheken oder anderen Fachportalen aufgebaut und betrieben werden, stellen einen wichtigen Wegweiser zu fachlich relevanten, qualitativ hochwertigen Informationsquellen im Internet dar. Aufgrund der ständigen Veränderungen, denen das Internet unterliegt, ist der Aufwand zur Pflege eines Fachinformationsführers von einer einzelnen Person – oft auch von einer einzelnen Institution – kaum zu leisten. Zur Unterstützung kooperativer Verfahren beim Betrieb von Fachinformationsführern steht als Ergebnis eines DFG Projekts mit DBClear nunmehr ein generisches Softwaresystem zur Nachnutzung zur Verfügung. DBClear zeichnet sich durch ein flexibles Metadaten-schema, umfangreiche Workflow- und Automatisierungsfunktionen und eine adaptierbare Internetoberfläche aus. Neben dem direkten Zugang zu den gespeicherten Inhalten durch das Suchen oder das Blättern in Klassifikationen kann der Zugriff auch über Z39.50 erfolgen.

Einleitung

Fachwissenschaftliche Informationsangebote im Internet wandeln sich mehr und mehr von Dienstleistungen, die von einer einzigen Person oder Institution erstellt werden, hin zu kooperativ von mehreren Partnern erstellten Produkten. Dies ist ein Prozess, der nicht nur von förderpolitischer Seite vorangetrieben wird (z.B. bei den Informationsverbünden des BMBF oder den Virtuellen Fachbibliotheken der DFG), sondern primär von den Anforderungen der Nutzer induziert wird. Diese verlangen nicht nur die Bündelung dezentraler Aktivitäten zur Bildung sowohl breit als auch tief gegliederter Fachportale, sondern auch die Verknüpfung der Ressourcen unterschiedlicher Fächer, um interdisziplinären Fragestellungen besser gerecht zu werden. Es ist klar, dass dies neben der Bereitschaft zur Zusammenarbeit auf der Seite der beteiligten Informationspro-

duzenten auch Softwaresysteme erfordert, die das räumlich und zeitlich verteilte Arbeiten unterstützen und aufgrund ihrer informationstechnologischen Architektur eine breite Nachnutzung erlauben. Von Entwicklungen im Bereich Digitaler Bibliotheken ist daher zu fordern, dass sie mehr als bisher die Aspekte des kooperativen Aufbaus von Informationssammlungen, des verteilten Arbeitens und der Kopplung dezentraler Dienste berücksichtigen. Dass Erfolge nur erzielt werden können, wenn derartige Systeme möglichst gut an den einzelnen Anwendungsfall angepasst werden können (z. B. Metadatenschemata, Arbeitsabläufe, benutzergruppen-spezifische Oberflächen) - und nicht andersherum - zeigt die Praxis.

Probleme beim kooperativen Aufbau von Fachinformationsführern

Fachinformationsführer, Clearinghouses oder Subject Gateways sind intellektuell aufgebaute Sammlungen von Verweisen auf fachlich relevante Quellen im Internet. Die Internetquellen werden dabei mittels Fachklassifikation oder auch regional gegliedert und mit einer Beschreibung versehen, die dem Nutzer die Beurteilung der Relevanz einer Quelle bezüglich seines Informationsbedarfes ermöglicht.

Die Auswahl der Quellen erfolgt in der Regel nach einer Reihe von Qualitätskriterien, die die inhaltliche Konsistenz auch über einen längeren Zeitraum hinweg sichern sollen (s. Bargheer 2003). Der Begriff „quality-controlled subject gateways“ (s. Koch 2000) findet für derartige Angebote häufig Anwendung, er geht allerdings über die reine Selektion von fachlich relevanten Quellen hinaus und umfasst auch Anforderungen an die Elemente, die zur Beschreibung und Strukturierung der Internetquellen verwendet werden (Metadaten) und an die fortlaufende Pflege des Angebots, zu der auch die regelmäßige Überprüfung der Aktualität der Beschreibungen und Erreichbarkeit der Internetquellen gehört.

Zur Beschreibung der Internetquellen werden unterschiedliche formale und inhaltliche Angaben verwendet, die sich vor allem an den fachlichen Anforderungen, an der Größe der Sammlung und an den verfügbaren personellen Ressourcen orientieren. Es existieren zwar erste Vorschläge zur Standardisierung von Metadaten für Fachinformationsführer (s. Becker et al. 2002), die großteils auf dem Dublin Core Metadata Element Set¹ (s. DCMI 2003) basieren und die Dewey Decimal Classification² zur Klassifikation vorschreiben, eine flächendeckende Akzeptanz scheint jedoch noch nicht in Sicht. Vor allem für umfangreiche existierende Sammlungen von Internetquellen stellt sich das Problem, bestehende Elemente auf den Standard abzubilden, gegenüber dem Standard feh-

1 <http://www.dublincore.org>

2 <http://www.oclc.org/dewey>

lende Elemente nachträglich zu erfassen oder für zusätzliche fachspezifische Elemente konsistente Erweiterungen zu definieren.

Sollen nun kooperativ von mehreren Personen – auch über Institutsgrenzen hinaus – Fachinformationsführer aufgebaut werden, ergeben sich sowohl positive Synergieeffekte als auch Hindernisse beim Zusammenführen von bestehenden Informationssammlungen. Inhaltliche und organisatorische Fragestellungen sind hier primär relevant.

Im inhaltlichen Bereich gilt es zunächst einmal, einheitliche Selektions- und Erschließungskriterien zu definieren. Dieser Prozess wird durch die Vorerfahrung der beteiligten Personen wesentlich beeinflusst, umfangreiche Vorarbeiten (z. B. Fachklassifikationen, Thesauri, Metadatenvorschläge bestehender Fachinformationsführer) aus unterschiedlichen Fachgebieten geben hierbei aber wertvolle Hilfestellung. Schwierigkeiten können auftreten, wenn bestehende Informationssammlungen integriert werden sollen. Bei unterschiedlicher Erschließungstiefe der Internetquellen ist zu entscheiden, welche Metadaten an den gemeinsamen Standard angepasst werden sollen. Der Verlust an Präzision oder Mehraufwand für die nachträgliche Erfassung sind die Folge. Der Anpassungsaufwand wird in der Praxis nicht nur durch die inhaltliche Diskussion und die Wünsche der Informationsnutzer bestimmt, sondern kann auch durch die feste Metadatenstruktur der zur verwendenden Software bestimmt werden. Im Extremfall können Einschränkungen bezüglich der Flexibilität des Metadaten-schemas den Einsatz einer spezifischen Software verhindern oder die kooperativen Bemühungen im Vorfeld scheitern lassen.

Auch im organisatorischen Bereich sind zunächst einige Hürden zu nehmen. Die Größe und Dynamik des abzudeckenden Fachgebiets bestimmt zunächst den personellen Aufwand zur Erfassung und Pflege der relevanten Internetquellen. Während neue Quellen zunächst noch relativ einfach mit Hilfe von Institutionsverzeichnissen und Internetsuchmaschinen gefunden werden können, erhöht sich der Aufwand zum Auffinden zusätzlicher relevanter Quellen mit steigender Größe der Sammlung. Dies betrifft in gleicher Weise die Überwachung der Aktualität oder Erreichbarkeit von Internetquellen. Soll der Aufwand auf mehrere Personen verteilt werden, zum Beispiel nach Fachgebiet oder Region, so sind Vorkehrungen notwendig, die Doppelarbeit vermeiden helfen, Arbeiten wenn möglich automatisieren und die Kommunikation zwischen den beteiligten Personen – den Workflow – unterstützen. Dies kann durch Hilfen bei der Erfassung oder Überwachung von Internetquellen geschehen, der automatischen Zuordnung von Quellen zu Redakteuren, oder die Aufteilung von Arbeitsschritten auf unterschiedliche Personen.

Das Projekt DBClear

Das Informationszentrum Sozialwissenschaften (IZ) in Bonn und die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) entwickelten in einem von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekt zwischen Oktober 2000 und September 2002 ein Softwaresystem zum Betrieb von Fachinformationsführern, das einer Vielzahl der oben angeführten Anforderungen gerecht wird. Ziel des Projekts war es, mit DBClear ein flexibles, generisches System zu entwickeln, das an beliebige Anwendungsfälle anpassbar ist. Es sollten nicht nur der SocioGuide des IZ und die Fachinformationsführer der SUB (s. Enderle et al. 1999) mit DBClear betrieben werden, sondern das System sollte als Standardsoftware für Digitale Bibliotheken und Virtuelle Fachbibliotheken nachnutzbar sein.

Zu Projektbeginn wurden neben dem SocioGuide und den Guides der SUB eine größere Zahl an anderen Portalen hinsichtlich Metadatenschema, Funktionalität und Technik analysiert. Die SUB und die Fachabteilungen des IZ erstellten daraus einen Anforderungskatalog hinsichtlich der Flexibilität des Metadatenschemas und Funktionen für das kooperative Arbeiten im Rahmen der Pflege eines Fachinformationsführers. Daneben entstand an der SUB eine Empfehlung für Qualitätskriterien zur Selektion fachlich relevanter Internetquellen (s. Bargheer 2003).

Technische Grundlagen

Die Auswahl von relevanten Technologien für die Realisierung von DBClear orientierte sich streng an der Anforderung, ein plattformunabhängiges, auf offenen Standards basierendes System zu entwickeln. Nach der Evaluierung kommerzieller Lösungen (insbesondere für die Workflowunterstützung), fiel die Entscheidung zugunsten einer Eigenimplementierung in Java, da nur damit der Betrieb bei möglichst vielen Institutionen – auch aus finanzieller Sicht – gesichert schien. Zum Betrieb von DBClear wird neben einer Java-Laufzeitumgebung eine sogenannte Servlet Engine zur Erzeugung dynamischer WWW-Seiten benötigt. Im Projekt kam hier Tomcat³ zum Einsatz, der sehr gut mit dem Apache WWW Server⁴ zusammenarbeitet und genau wie dieser kostenlos für eine Vielzahl von Betriebssystemen verfügbar ist.

Die Verwendung eines relationalen Datenbankmanagementsystems gewährleistet eine effiziente und konsistente Datenhaltung. Neben kommerziellen Produkten wie IBM DB2, Oracle, Microsoft SQL Server (mit alternativem JDBC-Treiber) wurden auch nicht-kommerzielle Systeme wie PostgreSQL erfolgreich

3 <http://jakarta.apache.org/tomcat>

4 <http://httpd.apache.org>

getestet. Auch mit dem multi-dimensionalen Datenbanksystem Caché wurden erste Tests erfolgreich durchgeführt. Für MySQL sind Tests geplant, sobald die aktuell in Entwicklung befindliche Version 4.1 allgemein verfügbar ist.

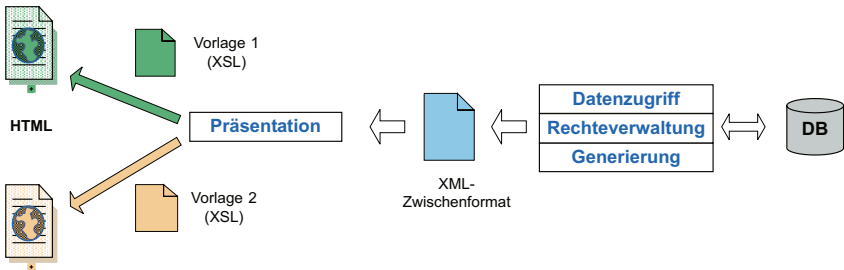


Abbildung1: DBClear Architektur

Die Architektur von DBClear (Abbildung 1) besteht aus mehreren Schichten mit spezifischen Aufgaben. Die Generierungsschicht liest für die Anzeige von Daten im Internet den Inhalt der Datenbank aus und generiert ein internes XML-Zwischenformat, welches alle relevanten Informationen enthält, die für die weitere Verarbeitung benötigt werden. Über die Rechteverwaltung können für einzelne Benutzergruppen spezifische Ansichten erzeugt werden, die sich in der Sichtbarkeit und Editierbarkeit von Metadatenelementen unterscheiden. Eine eigene Präsentationsschicht transformiert die XML-Daten dann nach HTML, so dass sie im Browser des Nutzers angezeigt werden können. Die Gestaltung dieser Darstellung erfolgt ausschließlich über XSL-Formatvorlagen, und ist von der Java Programmierung vollständig entkoppelt. Damit beschränken sich die Fertigkeiten, die zur Anpassung von DBClear an die eigenen Bedürfnisse erworben werden müssen, weitgehend auf die standardisierten XSL-Befehle. Zusätzlich wird es möglich, die einmal gespeicherten Daten in unterschiedlichen Kontexten oder abhängig vom Nutzer mit einem spezifischen Layout anzubieten.

Metadatenschema

Ein festes, nicht veränderliches Metadatenschema (z. B. Dublin Core oder das Vlib Application Profile) schied für DBClear aus, aufgrund der vielfältigen Metadatenschemata, die in existierenden Fachinformationsführern bereits verwendet werden. Vielmehr sollte die Migration von Altdaten durch die flexible Anpassbarkeit von DBClear auf die vorhandenen Datenstrukturen möglichst gut unterstützt werden. Ebenso sollte die Verwaltung mehrerer Sammlungen von

Internetquellen mit unterschiedlichen Metadatenschemata innerhalb einer DBClear-Installation ermöglicht werden. Ziel war nicht nur die Minimierung technischer Ressourcen, sondern die mögliche Integration heterogener Sammlungen von Internetquellen auf der Basis einzelner – eventuell zusätzlich zu erfassender – Metadatenelemente ohne eine vorherige Angleichung der Schemata.

Für DBClear wurde daher ein flexibles und dynamisches Modell für die Speicherung beliebiger Metadatenschemata entwickelt. Es basiert auf dem Konzept von Sammlungen von Internetquellen gleichen Typs (z. B. Homepages von Institutionen, Datenbanken im Internet, Lehrmaterialien), die *Bestände* genannt werden. Für jeden Bestand kann ein eigenes Metadatenschema definiert werden, das aus beliebigen *Attributen* mit kontrolliertem oder mit unkontrolliertem Inhalt bestehen kann. Attribute werden nur einmal definiert und können dann in allen Beständen einer DBClear-Installation genutzt werden können. Sofern zwei Bestände über gleiche Attribute verfügen (z. B. Klassifikation oder Titel), können ihre Inhalte über diese Attribute zusammen zur Navigation oder zur Suche angeboten werden.

Kontrollierte Attribute werden für Metadatenelemente verwendet, bei denen die möglichen Attributwerte a priori festgelegt werden sollen oder zueinander in hierarchischer Beziehung stehen (z.B. Schlagwörter, Klassifikationen, Thesauri, Categoriesystemen). Sie können entweder in DBClear importiert oder im Administrationswerkzeug definiert werden. Bei der Erschließung von Internetquellen stehen sie in Auswahllisten zur Verfügung oder können frei eingegeben werden. Beim Speichern einer Ressourcenbeschreibung findet ein Abgleich mit den erlaubten Werten statt; ungültige Werte werden zurückgewiesen.

Unkontrollierte Attribute werden für Metadatenelemente verwendet, bei denen unter Umständen keine (sinnvolle) Vorbelegung angegeben werden kann (z. B. bei Titel und Beschreibung einer Internetquelle oder bei Orts- und Personenamen). Bereits erfasste Attributwerte können allerdings bei der Neuerfassung als Vorlage dienen.

Die Abbildung 2 zeigt drei Bestände an Internetquellen (Projekte, Dissertationen und Institute), die jeweils aus unterschiedlichen kontrollierten und unkontrollierten Attributen zusammengesetzt sind. Für jedes Attribut kann angegeben werden, mit wie vielen Werten es minimal und maximal belegt werden darf (Kardinalität) und ob die Werte sprachabhängig sind. Abstracts oder Titel können somit in mehreren Sprachen parallel erfasst werden. Eigennamen oder andere Attribute, für die entweder eine Übersetzung nicht sinnvoll ist oder entsprechender Aufwand nicht geleistet werden kann, können als sprachunabhängig definiert werden.

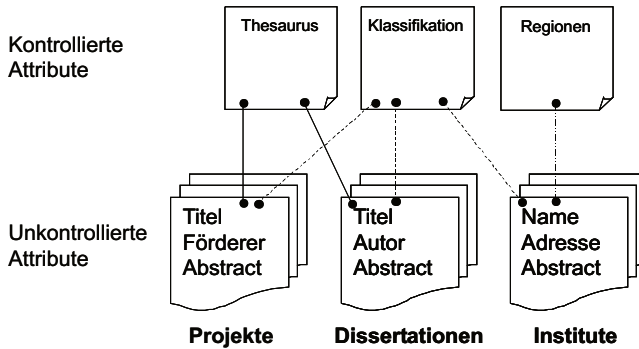


Abbildung 2: Metadatenbestände

Neben physikalisch vorhandenen Attributen sind auch so genannte *dynamische Attribute* möglich. Hierbei handelt es sich um kleine Java-Programme (Plug-Ins), die in DBClear über eine entsprechende Schnittstelle eingebunden werden. Dynamische Attribute können dabei zusätzliche Daten aus externen Datenquellen in DBClear verfügbar machen oder aus den gespeicherten Attributen neue erzeugen. Als Beispiel für das Einbinden einer externen Datenquelle kann eine Literatur- und Forschungsprojektdatenbank dienen, aus der die zu einer in DBClear beschriebenen Institution zugehörigen Publikationen und Forschungsprojektinformationen abgerufen werden, sobald ein Nutzer die Detailansicht der Institution öffnet. Ebenso lassen sich mit dynamischen Attributen virtuelle hierarchische Kategorien erzeugen, zum Beispiel eine Hierarchie Land – Stadt – Institutsart, die so nicht in DBClear gespeichert ist. Dabei werden während des Blätterns in der virtuellen Kategorie die für die drei Einzelattribute (Land, Stadt, Institutsart) vorhandenen Werteausprägungen analysiert, die Hierarchie aufgebaut und die entsprechenden Quellen dargestellt. Im angeführten Beispiel werden in einem ersten Schritt zu allen Internetressourcen die Länder ermittelt. Diese bilden die oberste Ebene des virtuellen Categoriesystems. Danach wird für jedes Land ermittelt, welche Städte den Ressourcen zugeordnet sind; diese werden dann dem Land als Unterkategorien beigelegt. Analog dazu wird für die Institutionsart vorgegangen. Zu jeder Stadt werden die zu ihren Internetressourcen vergebenen Institutionsarten ermittelt und als dritte Unterebene eingefügt, von der aus dann direkt auf die Internetressourcen zugegriffen werden kann.

Präsentation der Daten im Internet

Durch die strikte Trennung von Datenhaltung und Präsentation in DBClear kann das Erscheinungsbild im Internet frei an die Bedürfnisse der Nutzer angepasst werden. Als Technik kommt dabei die Transformation von XML-Daten anhand von Formatvorlagen (XSL) zum Einsatz. Als Basiszugang zu den Daten steht eine Such- und Blätterfunktion (Browsing) zur Verfügung.

GESIS Socioguide

country: language: fields of sciences: organization type: information on:

search terms: last modified: all Sort results by: last modified search reset

10 resources found

result page 1 1 for all entries

1. **Belarusian State University of Economics**, Belarus, Minsk
2. **Department of Marketing and Tourism**, University of Tirana, Faculty of Economics, Albania, Tirana
3. **Department of Finance and Accounting**, University of Tirana, Faculty of Economics, Albania, Tirana
4. **Department of Management**, University of Tirana, Faculty of Economics, Albania, Tirana
5. **Department of Economics**, University of Tirana, Faculty of Economics, Albania, Tirana
6. **Faculty of Economics**, University of Tirana, Albania, Tirana
7. **Center for Economic and Social Studies**, European University for Humanities (EHU), Belarus, Minsk
8. **Faculty of Economics**, Belarusian State University, Belarus, Minsk
9. **Faculty of Philosophy and Social Sciences**, Belarusian State University, Belarus, Minsk
10. **Centre for Sociological and Political Research**, Belarusian State University, Belarus, Minsk

If you feel like something is missing here, please use our [suggestion form](#) and help us complete our collection.

© GESIS-Aussenstelle, Berlin

Abbildung 3: Erweiterte Suche im SocioGuide

Die Abbildung 3 zeigt die erweiterte Suche im sozialwissenschaftlichen Fachinformationsführer SocioGuide des Informationszentrums Sozialwissenschaften⁵. Im oberen Teil befindet sich die Suchmaske, in die jedes Element des Metadatenschemas aufgenommen werden kann. Nach dem Abschicken der Suche findet sich darunter die Ergebnisliste. Die hier zu einer Internetquelle angezeigten Informationen können frei definiert und formatiert werden. Im SocioGuide öffnet das Anklicken eines Titels ein neues Browserfenster mit der Homepage der Internetquelle. Das gleiche Ziel kann durch klicken auf das Homepagesymbol am rechten Rand der Liste erreicht werden. Über das Infosymbol gelangt man zur Langdarstellung der Detailinformationen der Internetquelle, die im SocioGuide in einem eigenen Fenster erscheint. Als zusätzliche Funktionalität wurde eine ausführlichere Listendarstellung realisiert, die mit der Schaltfläche „i for all entries“ aktiviert werden kann. Es werden dann die ausführlichen In-

5 <http://www.gesis.org/SocioGuide/index.htm>

formationen zu allen Listeneinträgen eingeblendet, was den Ausdruck der Informationen erleichtert.

Neben der Suche kann in den Internetquellen auch geblättert (Browsing) werden. Im SocioGuide ist dies anhand der sozialwissenschaftlichen Klassifikation des Informationszentrums Sozialwissenschaften oder einer hierarchischen Liste der Regionen möglich (Abbildung 4). Das Blättern nach Regionen zeigt den Einsatz eines dynamischen Attributs, bei dem die tatsächlich bei der Erschließung vergebenen Werte dreier Attribute zu einem virtuellen Kategorienbaum verbunden werden.

GESIS Socioguide

The screenshot shows the GESIS Socioguide interface. At the top, there are tabs for 'simple search', 'advanced search', 'browse regions', and 'browse classification'. The 'browse regions' tab is active. Below the tabs, a breadcrumb trail shows 'Belarus > Minsk > other non university research institution'. The main content area is divided into two columns. The left column shows a hierarchical tree structure for browsing regions and institutions. The right column shows a list of results for the selected category.

Left Column (Hierarchy):

- Albania (0/19)
 - Belarus (3/27)
 - Minsk (0/24)
 - academy (4)
 - library, archive, information center (1)
 - other non university research institution (5)
 - public administration (1)
 - university, institute of higher education (13)
- Bulgaria (2/76)
 - Blagoevgrad (0/6)
 - international institution (1)
 - university, institute of higher education (5)
 - Botevgrad (0/1)
 - university, institute of higher education (1)
 - Bourgas (0/2)
 - university, institute of higher education (2)

Right Column (Results):

result page 1 (1 for all entries)

1. East Ukrainian Foundation for Social Research, Belarus, Minsk
2. Independent Institute for Socio-Economic and Political Studies, Belarus, Minsk
3. International Institute for Policy Studies (IIPS), Belarus, Minsk
4. International Institute of Labour and Social Relations, Belarus, Minsk
5. Minsk Center for Gender Studies, Belarus, Minsk

Abbildung 4: Browsing in einer virtuellen Kategorie

Unterstützung durch Automatisierung und Workflow

Ein Großteil des Arbeitsaufwands bei umfangreichen Sammlungen von Internetquellen fließt in die Pflege der bereits erschlossenen Quellen. Hier muss die Erreichbarkeit der Internetseiten überprüft werden und bei deren Änderung muss die inhaltliche Beschreibung aktualisiert und ergänzt werden. Der dabei zu leistende personelle Aufwand geht natürlicherweise für die Suche nach neuen relevanten Internetquellen verloren, so dass eine Unterstützung durch das Softwaresystem dringend notwendig ist. DBClear verfügt über Funktionen, die sowohl Routineaufgaben automatisieren können als auch die Verteilung von Arbeitsprozessen auf mehrere Personen ermöglichen. Da die Bearbeitung aller inhaltlichen Aspekte über einen Internetbrowser erfolgt, ist damit auch ortsunabhängiges Arbeiten möglich.

Bereits bei der Ersterfassung von Internetquellen setzt die Automatisierung von Arbeitsgängen an. Über frei definierbare Bedingungen (so genannte reguläre Ausdrücke) können Daten aus Internetseiten extrahiert und als Vorbelegung für Attribute verwendet werden. Standardmäßig erschließt DBClear aus Internetadresse, Zusatzinformationen des WWW-Servers und der eigentlichen Internetseite unter anderem Land, Sprache, Titel und Datum der letzten Änderung der Quelle und trägt dies in die dafür vorgesehenen Attribute ein. Zusätzlich wird über Internetsuchmaschinen die Anzahl der auf die Quelle verweisenden Internetseiten ermittelt (Backlinks), was als Kriterium für die Wichtigkeit und Qualität der Quelle verwendet werden kann.

Für die zur Erschließung einer Quelle notwendigen Arbeitsschritte kann ein Ablauf einzelner Tätigkeiten (Workflow) definiert und die einzelnen Schritte mehreren Personen oder Personengruppen zugeordnet werden. Anhand der Rechte der Personengruppen ergeben sich unterschiedliche Sichten auf die Daten, so dass jeweils nur bestimmte Attribute sichtbar oder editierbar sind. Am Ende jedes Arbeitsschritts besteht die Möglichkeit, den Status einer Internetquelle zu verändern (z. B. für Nutzer freigeben) oder an eine andere Person weiterzuleiten. Die Internetquelle erscheint dann mit einer entsprechenden Bemerkung oder einem Arbeitsauftrag (z. B. Quelle überprüfen) auf der ToDo-Liste der Person. Mit diesem Verfahren lassen sich auch inhaltliche oder regionale Bereiche bilden, die von unterschiedlichen Redakteuren bearbeitet werden. DBClear unterstützt gerade hier die Einbindung externer Experten über das Internet, so dass kooperativ umfangreiche und gleichzeitig tief erschlossene Fachinformationsführer aufgebaut werden können.

DBClear überprüft regelmäßig die Aktualität bzw. Erreichbarkeit der Internetquellen. Dabei kann sehr detailliert angegeben werden, wie häufig die Überprüfung stattfindet und welche Änderungen (z. B. Größenänderungen oder Änderungen der von der Seite wegführenden Links) zur automatischen Generierung eines Arbeitsauftrags an einen Redakteur führen.

Das Auffinden neuer Internetquellen wird von DBClear durch Suchprofile unterstützt, die für Internetsuchmaschinen (z. B. Altavista oder Google) definiert werden können. Diese Suchen werden in regelmäßigen Abständen ausgeführt und das Ergebnis mit den bereits dokumentierten Quellen verglichen. Bereits erfasste Quellen werden nicht mehr vorgelegt und nicht relevante Quellen können auf eine Negativliste (Blacklist) gesetzt werden, so dass sie zukünftig nicht mehr vorgeschlagen werden.

Schnittstellen zur Außenwelt

Eine flexible Export- und Importschnittstelle ermöglicht das Einfügen von Daten aus anderen Systemen und die Weitergabe von Daten. Dazu kann im Admi-

nistrationswerkzeug eine Abbildung der DBClear-Attribute auf die Felder des externen Formats vorgenommen werden. Damit ist zum Beispiel der Export von Daten im Dublin Core-Format möglich, auch wenn diese in DBClear in einem eigenen Metadatenschema gespeichert werden.

Zur Integration in andere Informationssysteme besitzt DBClear eine Z39.50-Schnittstelle, die auch in der Virtuellen Fachbibliothek Sozialwissenschaften⁶ (ViBSoz) für den GESIS SocioGuide verwendet wird. Der SocioGuide ist damit gemeinsam mit der Literaturdatenbank SOLIS und diversen Bibliothekskatalogen durchsuchbar.

Ausblick

DBClear steht nicht-kommerziellen Einrichtungen kostenlos zur Nachnutzung zur Verfügung. Die Pflege und Weiterentwicklung der Software wird zukünftig vom Informationszentrum Sozialwissenschaften gewährleistet. Interessierte Institutionen können im Rahmen einer einjährigen Projektverlängerung aktiv beim Einsatz von DBClear unterstützt werden. Die Hilfestellung reicht dabei von der Installation über die Migration von Altdatenbeständen und Unterstützung bei der Konfiguration der Benutzungsoberfläche bis zur funktionalen Erweiterung der Software in begrenztem Maße.

Darüber hinaus sind für 2003 die Implementierung der Schnittstelle zum in Entstehung befindlichen deutschen Wissenschaftsportal VASCODA⁷ und eines Moduls für Veranstaltungskalender vorgesehen. Des Weiteren befindet sich eine dynamische Schnittstelle in Planung, über die – zusammen mit einer Heterogenitätskomponente – die in DBClear verwendeten Formate und Erschließungssysteme auf beliebige andere abgebildet werden können (z. B. DC und DDC „on-the-fly“).

DBClear wird bereits von der Arbeitsstelle Friedensforschung Bonn⁸ (AFB) für das AFB-Forschungsregister und den SocioGuide des Informationszentrum Sozialwissenschaften eingesetzt. Zur Zeit befindet sich der Fachinformationsführer der Virtuellen Fachbibliothek Gegenwartkunst⁹ und die Sammlung des Osteuropa-Netzwerks¹⁰ im Aufbau. Mehr als zehn weitere Institutionen haben bereits schriftlich ihr Interesse an einer Nachnutzung von DBClear bekundet.

6 <http://www.vibsoz.de>

7 <http://www.vascoda.de>

8 <http://www.priub.org>

9 <http://vifaart.slub-dresden.de/>

10 <http://www.osteuropa-netzwerk.de/>

Literatur

- Bargheer, Margo (2003): Qualitätskriterien und Evaluierungswege für wissenschaftliche Internetressourcen - Ein Report für die bibliothekarische und dokumentarische Praxis. Niedersächsische Staats- und Universitätsbibliothek Göttingen, 1. Auflage, Göttingen 2002.
(<http://webdoc.sub.gwdg.de/ebook/aw/2003/bargheer/v10.pdf>)
- Becker, Hans J.; Hengel, Christel; Neuroth, Heike; Weiss, Berthold; Wessel, Carola (2002): Die Virtuelle Fachbibliothek als Schnittstelle für eine fachübergreifende Suche in den einzelnen Virtuellen Fachbibliotheken: Definition eines Metadaten-Kernsets (VLib Application Profile). In: Bibliotheksdienst 36 (2002), Nr. 1, S. 35 - 51.
- DCMI (2003): Dublin Core Metadata Element Set, Version 1.1: Reference Description. Dublin Core Metadata Initiative
(<http://dublincore.org/documents/2003/02/04/dces/>).
- Enderle, W. et al. (1999): Das Sondersammelgebiets-Fachinformationsprojekt (SSG-FI) der Niedersächsischen Staats- und Universitätsbibliothek Göttingen: GeoGuide, MathGuide, Anglo-American History Guide und Anglo-American Literature Guide. dbi-materialien 185, DBI Berlin.
- Koch, Traugott (2000). Quality-controlled subject gateways: definitions, typologies, empirical overview. In: Subject gateways, Special issue of „Online Information Review“, Vol. 24:1, 2000, pp.24 - 34.
(Manuskript: <http://www.lub.lu.se/~traugott/OIR-SBIG.txt>)

PhysNet und seine Spiegel - Das Projekt SINN

Michael Hohlfeld, Thomas Severiens

Institute for Science Networking Oldenburg GmbH

Abstract

The aim of the project SINN - Suchmaschinennetzwerk im Internationalen Naturwissenschaftlichen Netz - is to enhance the distributed information system PhysNet (www.physnet.net) to a fast, secure and scalable service by setting up a network of independent but cooperating Harvest-brokers and Harvest-gatherers and building an active network of PhysNet-mirrors. We will report on the current development status of this mirror-network and about its usage.

Einführung

Der fachspezifische Informationsdienst PhysNet (www.physnet.net) wird seit vielen Jahren im Auftrage der Europäischen Physikalischen Gesellschaft und verschiedener nationaler Fachgesellschaften durch die Institute for Science Networking Oldenburg GmbH (kurz ISN) entwickelt und zusammen mit einer Vielzahl internationaler Partner gemeinschaftlich betrieben und gepflegt.

Der PhysNet-Dienst bietet verschiedene thematische Linklisten, beispielsweise Links zu Physik-Fachbereichen, Instituten und Institutionen weltweit, zu wissenschaftlichen Dokumenten bzw. Dokumentenquellen (wie Publikationslisten, PrePrints und so genannter grauer Literatur), aber auch zu physikalischen Zeitschriften, Konferenzen, Stellenmärkten, Lehr- und Lernmaterialien und anderen relevanten Informationsquellen. Ergänzt wird der Dienst durch mehrere Suchmaschinen, welche auf dem Harvest-System basieren, und welche es z.B. ermöglichen, nach verschiedenen Informationen oder Publikationen auf den in PhysNet gelisteten Web-Servern zu suchen.

In diesem Beitrag soll auf einige aktuelle Entwicklungen bei PhysNet eingegangen werden, welche im Rahmen des Drittmittelvorhabens SINN - Ein Suchmaschinennetzwerk im Internationalen Naturwissenschaftlichen Netz - entwickelt und umgesetzt worden sind.

SINN ist ein Vorhaben im Rahmen des DFN-Projektes „Einsatz von Netzdiensten im Wissenschaftlichen Informationswesen“, und wird mit Mitteln des Bundesministeriums für Bildung und Forschung (bmb+f) und des Landes Nie-

dersachsen gefördert. Die ISN betreibt dieses Projekt in enger Kooperation mit mehreren der internationalen Partnern des PhysNet-Dienstes.

Begonnen wurde mit dem Projekt im Februar 2001, auslaufen wird es im Oktober 2003. Weitere Informationen zum SINN-Projekt finden sich auf der Projekt-Homepage unter <http://isn-oldenburg.de/projects/SINN>.

Was macht SINN für PhysNet?

PhysNet ist ein verteilter Informationsdienst, der nicht nur weltweit verteilte Informationen der Physik gebündelt zugänglich macht, sondern dessen Betrieb und Pflege ebenfalls global verteilt organisiert ist. Diese 'Arbeitsteilung' zu intensivieren und gleichzeitig den Dienst hinsichtlich der Nutzerfreundlichkeit zu optimieren ist das Ziel und die Herausforderung des Projektes SINN.

Eine der Anforderungen die an einen aus Nutzersicht optimalen (verteilten) Informationsdienst gestellt werden müssen, ist die Schnelligkeit und Stabilität des Dienstes. Im Rahmen des SINN-Projektes soll PhysNet daher durch den Aufbau eines aktiven Spiegelnetzwerkes, also durch das Spiegeln des Dienstes auf weltweit verteilte Server zu einem in der Nutzung belastbaren, schnellen und sicheren Dienst ausgebaut werden. Anfragen an das System sollen immer schnell und stets bearbeitet werden können, unabhängig vom Standort des Nutzers, von der Netzbelastung oder dem Ausfall von Programmen oder Hardware.

Der Aufbau eines solchen Spiegel-Systems beinhaltet aber nicht nur das Spiegeln der HTML-Seiten, d.h. der Linklisten usw., sondern auch und insbesondere das Spiegeln der Suchmaschinen des PhysNet-Dienstes. Durch den Aufbau eines kohärent gekoppelten Netzes von Harvest-Brokern und Gatherern soll erreicht werden, dass das PhysNet-System mehrere global verteilte Spiegel des vollen Datensatzes von weltweit regional erstellten Index-Dateien vorhält. Der Nutzer des Systems soll dann z.B. bei einem Ausfall eines (oder mehrerer) der beteiligten Suchmaschinen-Interfaces automatisch auf einen anderen, funktionierenden Spiegel geleitet werden.

Der Aufbau eines solchen verteilten Suchmaschinennetzwerkes stellt dabei eine gewisse Herausforderung dar, denn für den notwendigen Austausch technisch-administrativer Informationen zwischen den verschiedenen Informationsbrokern des verteilten PhysNet-Suchmaschinensystems und für die Query-basierte Vernetzung (also den Anfragenaustausch) werden standardisierte Protokolle bzw. Sprachen benötigt. Eine wesentliche Aufgabe des Projektes ist daher ein entsprechendes (XML-basiertes) Protokoll und eine (XML-Query basierte) Sprache zum Anfragenaustausch zu entwickeln und in das System zu implementieren.

Internationale Gremienarbeit

Aufgebaut werden soll(te) dabei wesentlich auf die Entwicklungen der W3C-Working group XML-Query, welche ein solches Protokoll als Standard definieren soll. Natürlich sind auch die Aktivitäten der Open Archive Initiative, bei der Metadaten über Dokumente zwischen Archiv-Servern mittels eines XML-Protokolls ausgetauscht werden, für das Projekt von Interesse. Die ISN ist hier mit einem anderen Projekt - „Open Archives: Distributed services for physicists and graduate students (OAD)“ - direkt in konkrete Entwicklungen eingebunden, so dass das Projekt von Synergieeffekten profitiert. XML-Query geht im Vergleich aber weiter als OAi und definiert insbesondere auch das Weiterreichen von Suchanfragen und das Zurückgeben der Antworten zwischen Suchmaschinen.

Zu Projektbeginn wurde davon ausgegangen, dass XML-Query relativ schnell, d.h. in der ersten Hälfte der Projektlaufzeit, den Status einer Accepted Recommendation des W3C erreichen würde. Bei der intensiven Einarbeitung in die Thematik und aktuellen Diskussionen der XML-Query-Entwicklergruppe zeichnete sich dann aber ab, dass dies so schnell nicht der Fall sein würde. Die im Projekt vorgesehenen Entwicklungen und Implementierungen wurden daher zugunsten einer aktiveren Mitarbeit in der XML-Query Entwicklergruppe zurückgestellt, die ohnehin im Projekt vorgesehene internationale Gremienarbeit intensiviert. Die Mitgliedschaft des DFN-Vereines beim W3C wurde dazu genutzt, offizielles Mitglied der XML-Query-Entwicklergruppe zu werden.

Die Arbeit in diesem Gremium ist sehr arbeitsintensiv, sie wird durch einen sehr intensiven Email-Austausch, wöchentliche Telefonkonferenzen und die Teilnahme an den Face-to-Face-Meetings der Gruppe bestimmt. Durch die aktive Teilnahme hat das Projekt SINN aber sehr profitiert, denn zum einen war und ist es nun möglich, auf bestimmte Richtungsentscheidungen zugunsten konkreter Anforderungen aus den Naturwissenschaften und deren Informationsdiensten (wie dem PhysNet-Dienst) Einfluss zu nehmen. Zum anderen konnten auf diese Weise schon eine ganze Reihe von konkreten Vorarbeiten für die Umsetzung innerhalb des SINN-Projektes geleistet werden, welche sich stark an den zu erwartenden Standards orientieren.

Die Diskussionen innerhalb der Gruppe sind mittlerweile weit fortgeschritten. Noch am Ende des letzten Jahres wurde daher davon ausgegangen, dass das XML-Query auf der nächsten W3C-Konferenz, welche vom 20. bis 24. Mai 2003 in Budapest stattfindet, die Protokollversion 1.0 zur Abstimmung stellt. Inzwischen ist aber klar, dass auch dieser Zeitplan nicht eingehalten werden kann, XML-Query wird erst deutlich nach Budapest - in der zweiten Jahreshälfte - zum Standard (Accepted Recommendation) werden. Dieser Standard ist dann verbindlich für alle XML-Repositories und deren Inhalte, welche an dem Aufbau eines XML-Query-basierten Netzwerkes teilnehmen wollen.

Das PhysNet-Spiegel-System

Der Aufbau des PhysNet-Spiegel-Systems erfolgt(e) schrittweise. Zu Beginn des Projektes wurde eine Spiegelsoftware entwickelt, welche es erlaubt, den PhysNet-Dienst auf anderen Webservern zu spiegeln. Dies beinhaltete zunächst die reinen HTML-Seiten (also die Rahmendateien, Bilder und Linklisten), mittlerweile aber auch die Konfigurationsdateien für die verschiedenen Harvest-Gatherer der PhysNet-Suchmaschinen.

Derzeit gibt es Spiegel an 13 Standorten weltweit, den „Root-Server“ in Oldenburg mit inbegriffen. Diese finden sich in Deutschland, Großbritannien, der Schweiz, Russland, Polen, Frankreich, den Vereinigten Staaten, Tschechien, Ungarn, den Niederlanden, Österreich und Indien.

Im Wesentlichen aus dem oben genannten Grund - dem bis dato fehlenden Standard - konnte das geplante Suchmaschinennetzwerk innerhalb des Spiegel-Systems bisher nicht umgesetzt werden. Auf einen weiteren Grund und wie es nun pragmatisch aufgebaut werden soll, wird weiter unten im Abschnitt ‘Einbindung internationaler Partner’ kurz eingegangen.

Der PhysNet-Kern

Wesentlicher Bestandteil des PhysNet-Spiegel-Systems ist der so genannte PhysNet-Kern, welcher seit Ende 2002 unter der zentralen URL www.physnet.net zu finden ist.

Ein Programm überprüft hier automatisch und in kurzen Abständen (alle 5 Minuten), ob jeder der PhysNet-Spiegel-Server erreichbar ist und ob die dort gespiegelten Daten aktuell, d.h. nicht älter als 14 Tage sind. Dem Besucher dieses PhysNet-Kernes wird eine Liste der verfügbaren und aktuellen Spiegelserver angezeigt. Die Reihenfolge hängt dabei vom Ort des Nutzers und der Antwort-Geschwindigkeit des Spiegelservers ab, so dass der vermutlich für den Nutzer schnellste Spiegelserver als erstes gelistet wird. Nach etwa 10 Sekunden wird der Nutzer dann auch automatisch auf diesen PhysNet-Spiegel weitergeleitet.

Bei der Feststellung dieses vermutlich schnellsten Spiegel-Servers findet bereits ein Teil jener Programme eine erste Anwendung, die im Projekt für die Verteilung der Suchanfragen an die Broker im Suchmaschinennetz programmiert wurden.

Ergänzend zu diesem PhysNet-Kern wird den Nutzern auch auf allen Seiten des PhysNet-Dienstes angezeigt, wo sich vermutlich sein nächstgelegener PhysNet-Spiegel befindet, sofern er diesen nicht schon nutzt.

Nutzung und Nutzeranalyse

Ein weiterer Schwerpunkt im SINN-Projekt ist die Nutzeranalyse durch die Auswertung von Nutzungsstatistiken und die Verbesserung der Nutzerakzeptanz durch gezielte Aktionen.

Abbildung 1 zeigt die Gesamtnutzung des PhysNet Dienstes über alle Spiegelstandorte und Teildienste summiert, wobei hier die tägliche Nutzung (Klicks/ Hits und Visits), jeweils gemittelt über eine Woche aufgezeigt werden.

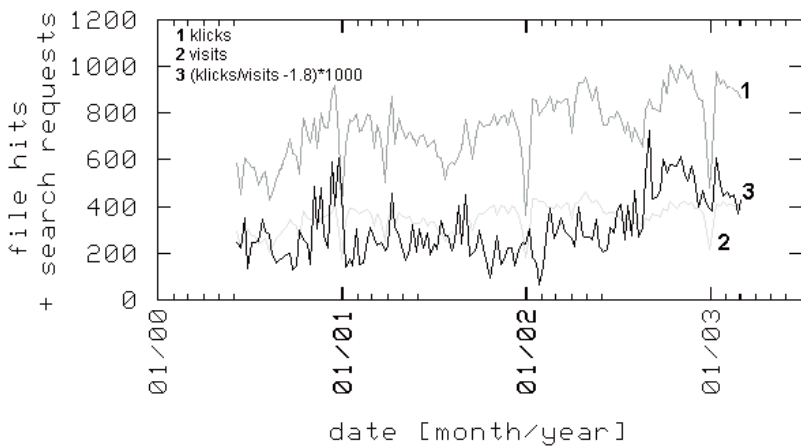


Abbildung 1: Gesamtnutzung des PhysNet Dienstes.

Dargestellt ist die durchschnittliche tägliche Nutzung (Klicks und Visits)

Ermittelt wird die Gesamtnutzung durch versteckte 1 Pixel große Bilder auf den verschiedenen Webseiten von PhysNet, welche unabhängig vom jeweiligen Spiegel-Server immer vom Oldenburger PhysNet-Server geladen werden. Der Aufruf dieser Bilder wird in einem Apache-Logfile protokolliert und ermöglicht so auch eine Aufschlüsselung bezüglich der Nutzung der einzelnen Teildienste.

Interessant ist das Surf- und Search-Verhalten der Nutzer. Wie das Verhältnis von Klicks zu Visits in der Abbildung zeigt, klickt jeder Nutzer im Durchschnitt zweimal, wechselt also z.B. von der Startseite in einen Teildienst und wird dann auf einen anderen Server geleitet. Seit Beginn des Projektes steigt dieser Wert aber ganz allmählich.

Aus den Log-Files lässt sich allerdings auch ermitteln, dass nach wie vor nur sehr wenige Nutzer des PhysNet-Dienstes die Suchmaschinen benutzen (ca.

10%). Hier bestand schon vor Projektbeginn und besteht weiterhin Handlungsbedarf, die Suchmaschinen den Nutzern nahe zubringen.

Optimierung der Suchmaschinen

Eine Tätigkeit im Projekt war und ist daher auch die Optimierung der in PhysNet angebotenen Suchmaschinen insbesondere der Harvest-Broker. Diese lieferten bisher bei einer Recherche unsortierte Listen von Suchergebnissen. Die Resultate führten den Nutzer zwar in der Regel auf Seiten oder Dokumente, auf denen der Suchbegriff auch zu finden war, eine Gewichtung der Ergebnisse fand aber nicht statt, was aus Nutzersicht natürlich nicht zufrieden stellend ist.

Die PhysNet-Suchmaschinen wurden daher mit einer Ranking-Funktion versehen. Hierbei werden Dokumente, die mit Metadaten versehen sind und in diesen das Suchwort enthalten, mit einem höheren Gewicht versehen. Dublin-Core-Metadaten werden dabei höher gewichtet. Die Suchergebnisse werden zudem nicht mehr in einer langen Liste angezeigt, sondern, wie auch bei anderen Suchdiensten üblich, in einzelne Ergebnisseiten aufgeteilt, was dem Nutzer sehr entgegen kommt. Auch wird das Suchwort in der Ergebnisanzeige hervorgehoben. Das Ranking verbessert somit wesentlich die Ergonomie der Recherche-Möglichkeiten in PhysNet.

Bei der Entwicklung der entsprechenden Ranking-Algorithmen konnte auf Ergebnisse des CARMEN-Projektes zurückgegriffen werden, in dem eine Verknüpfung der Dokumentendienste PhysDoc und MPRESS implementiert wurde, u.a. mit der Möglichkeit, über beide Dienste parallel zu suchen. Die hier verwendeten Algorithmen konnten den Bedürfnissen der anderen PhysNet-Suchmaschinen angepasst werden.

Nutzung der PhysNet-Spiegel

Im Folgenden soll die Nutzung der einzelnen PhysNet-Spiegel etwas detaillierter betrachtet werden.

Der Aufruf der schon erwähnten versteckten Bilder wurde durch einen HTTP-Get-Request ergänzt, dessen Variable mittels ServerSideIncludes vom ausliefernden Server auf dessen Hostnamen gesetzt wird. Auf diese Weise kann die Gesamtstatistik nach den Requests sortiert und anschließend ein normales Apache-Log-File für jeden der Spiegel-Server getrennt angelegt und ausgewertet werden.

Tabelle 1 zeigt die Nutzungsverteilung für den Monat Februar. Zu diesem Zeitpunkt sind bereits 27,5% der Nutzung auf die Spiegel des PhysNet-Dienstes verteilt. Seit Beginn des Projektes steigt diese Zahl stetig an. Wie positiv diese

Nutzungsverteilung einzuschätzen ist, zeigt der Vergleich mit dem bekannten zentralen Preprint-Dienst ArXiv.org, dessen global verteilte Spiegel nur von ca. 10% der Nutzer besucht werden.

Tabelle 1: Prozentuale Nutzung der PhysNet-Spiegel und regionale Nutzerverteilung (Unr.. = unresolved/unknown)

PhysNet-Spiegel	Nutzung (%)	Nutzung nach Herkunft für Februar 2003 (%)							
		1		2		3		4	
de.physnet.net (vorher www.physics-network.org)	72,5	24,4	Unr.	19,1	.de	12,7	.net	5,5	.it
physnet.eprints.org	9,1	25,7	Unr.	9,3	.net	9,2	.com	6,6	.de
www.eps.org/PhysNet	6,0	25,5	Unr.	14,6	.com	11,5	.pl	8,8	.net
www.physnet.ru	2,8	34,3	.ru	22,2	Unr.	18,1	.net	3,6	.de
pl.physnet.net	2,0	79,9	.pl	13,2	Unr.	2,3	.com	1,3	.net
www.ccsd.cnrs.fr/PhysNet	1,8	49,0	.fr	25,5	Unr.	4,9	.net	3,9	.com
www.phys.vt.edu/PhysNet	1,6	24,7	.edu	27,4	Unr.	18,3	.net	8,2	.com
www.uni-bayreuth.de/PhysNet	1,1	31,5	.de	30,2	.net	23,9	Unr.	2,7	.com
cz.PhysNet.net	0,9	55,1	.cz	11,9	.sk	16,5	Unr.	4,6	.com
hu.physnet.net	0,6	33,8	.ch	27,6	.hu	12,3	Unr.	9,8	.net
www.nikhef.nl/PhysNet	0,5	34,0	.nl	15,7	Unr.	15,0	.fr	7,8	.net
in.physnet.net	0,2	54,4	Unr.	10,9	.edu	10,9	.net	4,4	.ch
at.physnet.net	0,2	50,0	.at	13,9	Unr.	9,7	.fr	6,9	.org
PhysNet gesamt		22,4	Unr.	20,9	.de	11,0	.net	5,1	.com

Die regionale Verteilung der Nutzer lässt sich in der Regel ebenfalls auflösen. Tabelle 1 zeigt auch die Herkunftsländer der jeweils ersten vier „Hauptnutzer“ der verschiedenen Spiegel in Prozent für den Monat Februar 2003.

Einschränkend muss erwähnt werden, dass die Datenbasis für die weniger genutzten Spiegel relativ dünn ist und somit eine gewisse Verzerrung der Nutzerverteilung stattfindet. Die Hauptnutzung wird natürlich stark durch die Nutzung direkt am Spiegelstandort mitbestimmt, da dort der Spiegel bekannt ist/sein sollte (dieses lässt sich zumindestens teilweise aus den Statistiken erkennen).

Insgesamt lassen sich aber durchaus Unterschiede in der Nutzerverteilung und eine Tendenz feststellen. Während nämlich die Spiegel, bei denen in der

URL mit dem Countrycode das Land des Spiegelstandortes eindeutig zu identifizieren ist, die Hauptnutzung eben aus diesem Land erfahren, ist dies bei den Spiegeln mit den .org-Adressen anders. Dort ist die regionale Verteilung der Nutzer viel unspezifischer.

Aus dieser Erkenntnis resultiert daher auch der Ansatz Sub-Domains der Form cc.physnet.net zu vergeben, wobei cc eben für die Länderkennung (Countrycode) steht. Die einzelnen Spiegel werden seit Ende 2002 nun nach und nach auf diese Namens-Struktur umgestellt.

Dort wo diese Umstellung schon vollzogen wurde, war im Übrigen auch vorher schon durch die Länderkennung der anbietenden Institution der Standort des jeweiligen Spiegels erkennbar. Eine Ausnahme bildet hier der Haupt-Server in Oldenburg, der längere Zeit unter www.physics-networks.org propagiert wurde. Relativ stark verbreitet ist allerdings auch noch die Verwendung der ursprünglichen URL <http://physnet.uni-oldenburg.de/PhysNet/>.

Mit der Analyse der Nutzungsverhaltens wird im weiteren Projektverlauf fortgefahren, so dass zum Projektabschluss eine ausführliche Auswertung vorgelegt werden kann.

Zur Verbesserung der Nutzerakzeptanz und Steigerung des Bekanntheitsgrades der einzelnen Spiegel sollen (und müssen) weitere Aktionen durchgeführt werden. Als zumindestens kurzfristig sehr effektiv haben sich dabei immer zielgerichtete Mailaktionen erwiesen. Diese werden daher in Zusammenarbeit mit den beteiligten internationalen Partnern in nächster Zeit forciert werden. Auch sollen zukünftig Infotexte in verschiedenen Sprachen angeboten werden.

Einbindung der internationalen Partner

Die direkte Einbindung der internationalen Partner des PhysNet-Dienstes ist eine weitere Ebene der Tätigkeiten im SINN-Projekt.

Hierfür war im Projekt die Durchführung mehrerer technischer Workshops vorgesehen, zu denen die Partner nach Oldenburg eingeladen wurden, um gemeinsam das Vorgehen zu diskutieren und zu vereinbaren. Mit diesen Treffen wurden auch neue Partner für den PhysNet-Dienst und SINN gefunden. Der erste Workshop SINN01 fand im Dezember 2001 statt, der zweite (SINN02) dann im November 2002, jeweils begleitet von einer eintägigen Konferenz zur Thematik.

Hauptthema der letzten Tagung (SINN02), war das noch aufzubauende Suchmaschinensystem innerhalb des PhysNet-Spiegel-Systemes, für dessen Realisation bisher nicht nur internationale Standards fehlen. Es wurde pragmatisch daran gearbeitet, wie zukünftig mehr Spiegelstandorte eine eigene Suchmaschine anbieten und wie diese dann in das System eingebunden werden können. In der Tat ist es nämlich so, dass bisher nur ein sehr kleiner Teil der Partner,

die einen Spiegel des Portals betreiben auch eine eigene Suchmaschine basierend auf der Harvest-Software installiert haben. Die Gründe hierfür sind lokal durchaus verschieden (keine Personalkapazität, keine ausreichende Hardware, keine Kenntnisse bzw. eine gewisse Hemmschwelle sich mit Harvest etc. auseinanderzusetzen, ...) und spiegeln die organisatorische Herausforderung beim Betrieb eines verteilten Informationsdienstes mit verteilter Arbeitskraft wieder.

Zukünftig soll prinzipiell die Möglichkeit bestehen, lokal/regional auch andere offene Systeme als das Harvest-System zu nutzen, deren Suchinterfaces dann direkt in die jeweiligen Linklisten in PhysNet eingebunden werden. Auf diese Weise soll die Anzahl der lokalen Suchmaschinen deutlich steigen, so dass die gewünschte Vernetzung der Suchmaschinen zu einem redundanten, ausfallsicheren Netzwerk erreicht werden kann.

Generell soll aber weiterhin die Harvest-Software mit ihren Funktionalitäten genutzt und zukünftig in engerer Zusammenarbeit mit den derzeitigen Entwicklern auch neuere Versionen erprobt und ggfs. implementiert werden. Hierzu beigetragen haben sicherlich auch die sehr fruchtbaren Diskussionen mit einem der derzeitigen Entwickler von Harvest auf der SINN03-Tagung.

Um die internationale Zusammenarbeit im Rahmen des PhysNet-Dienstes weiter zu intensivieren, sollen zukünftig weitere Veranstaltungen dieser Art durchgeführt werden. Als Abschluss des SINN-Projektes wird vom 17. bis zum 19. September 2003 die Konferenz SINN03 mit dem Titel „Worldwide Coherent Workforce, Satisfied Users - New Services for Scientific Information - “ in Oldenburg stattfinden. Thematischer Schwerpunkt dieser Tagung wird sein, wie die Arbeit innerhalb verteilter Informationsdienste einerseits verteilt, andererseits gebündelt organisiert werden kann und welche neuen Dienste und Angebote zukünftig nötig sein werden, um einen aus Nutzersicht optimalen Informationsdienst aufzubauen.

Weitere Informationen finden sie auf der Tagungs-Homepage unter <http://www.isn-oldenburg.de/projects/SINN/sinn03/index.html>.

Zusammenfassung und Ausblick

In diesem Beitrag wurden aktuelle Tätigkeiten rund um den verteilten Informationsdienst PhysNet im Rahmen des Projektes SINN vorgestellt. Am Ende dieses Projektes soll PhysNet zu einem Informationssystem für die Physik geworden sein, das die gebündelte Information möglichst vieler fachspezifischer Server vorhält, über verteilte Suchmaschinen zugänglich macht und eine einheitliche Nutzeroberfläche zu möglichst vielen fachspezifischen Datenquellen bietet.

Der Nutzer dieses Systems braucht sich im Prinzip nur die URL www.physnet.net merken und wird automatisch auf den am schnellsten zu erreichenden und funktionierenden PhysNet-Spiegel geleitet.

Eine der wesentlichen Aktivitäten im SINN-Projekt, welche auch zukünftig fortgesetzt und intensiviert werden muss, ist und war die Arbeit in internationalen Standardisierungsgremien, hier speziell die aktive Teilnahme in der XML-Query-Entwicklergruppe des W3C. Das Projekt hat hiervon sehr profitiert. Insbesondere konnten viele Vorarbeiten zu den anstehenden Entwicklungen im Rahmen von SINN nahe am Entwicklungsstand von XML-Query getätigt werden. Inzwischen ist es absehbar, wann XML-Query zum internationalen Standard werden wird und dass somit die im Projekt geplanten Implementationen eines XML-basierten Protokolls zum Austausch technisch-administrativer Informationen und eine XML-Query basierte Sprache zum Abfragen austausch standardkonform abgeschlossen werden können.

Die Optimierung der PhysNet-Suchmaschinen durch eine bessere Recherche-Möglichkeit, diene der Verbesserung der Nutzerakzeptanz. Zu einer wesentlichen Mehrnutzung der Suchmaschinen hat dieses allerdings noch nicht geführt. Hier besteht also weiterhin Handlungsbedarf in der Form, den Nutzern des Systemes die Suchmaschinen näher zu bringen. Ein wesentlicher Schritt hierzu ist, nun an allen Spiegelstandorten lokal betriebene Suchmaschinen zu installieren, diese in das System zu integrieren, um so die Vernetzung der Suchmaschinen zu einem redundanten, ausfallsicheren Netzwerk zu erreichen.

Die Analyse des Nutzungsverhaltens hat deutlich gezeigt, dass die weltweit verteilten Spiegel des PhysNet-Dienstes in den einzelnen Ländern angenommen und genutzt werden; durch den PhysNet-Kern wird dieses sicher weiter forciert werden, da die Nutzer entsprechend 'verteilt' werden.

Die fortlaufende Nutzeranalyse soll zudem zeigen, wie sich die Auswertung des Nutzerverhaltens zur Optimierung der Nutzerakzeptanz eines Dienstes in der Wissenschaft nutzen lässt. Dies wird das bestimmende Thema der das Projekt abschließenden SINN03-Konferenz im September 2003.

Literatur/Quellen

- T. Severiens: PhysNet-Searchengines: Implementation of Ranking. Vortrag auf der Tagung SINN02: 2nd International Technical Workshop of the project SINN, Conference on Open Distributed Science Information Management, Oldenburg, 6 - 8 Nov. 2002
http://isn-oldenburg.de/projects/SINN/sinn02/severiens_harvest-ranking_talk.html
- T. Severiens: XML-Query. Vortrag auf der Tagung SINN02: 2nd International Technical Workshop of the project SINN, Conference on Open Distributed

- Science Information Management, Oldenburg, 6 - 8 Nov. 2002
http://isn-oldenburg.de/projects/SINN/sinn02/severiens_xml-query_talk.html
- M. Hohlfeld: Status of SINN. Vortrag auf der Tagung SINN02: 2nd International Technical Workshop of the project SINN, Conference on Open Distributed Science Information Management, Oldenburg, 6 - 8 Nov. 2002
http://isn-oldenburg.de/projects/SINN/sinn02/hohlfeld_status_talk.html
- M. Hohlfeld, T. Severiens: PhysNet - Distributed Online Information Services for Physics. Vortrag beim ftw. Forschungszentrum Telekommunikation Wien, Vienna, Austria, 27 Sep. 2002 und am KFKI Research Institute for Particle and Nuclear Physics, Budapest, Hungary, 24 Sep. 2002
http://www.isn-oldenburg.de/talks/budapest_sept2002/
- M. Hohlfeld, T. Severiens: Das Projekt SINN: Ein Suchmaschinennetzwerk im Internationalen Naturwissenschaftlichen Netz. Vortrag auf der 24. DGI-Online-Tagung: Content in Context - Perspektiven der Informationsdienstleistung, Frankfurt am Main, 3. - 5. Juni 2002
<<http://www.isn-oldenburg.de/talks/dgi-online2002/>>
Veröffentlichung im Tagungsband der Konferenz [ISSN 1438-9274]
<<http://www.isn-oldenburg.de/~hohlfeld/dgi-online2002.html>>
- M. Hohlfeld: Vernetzung von verteilten Informationssystemen. Vortrag auf der Konferenz '8th Annual Meeting of the IuK Initiative: Offene Systeme für die Kommunikation in Wissenschaft und Forschung', Ulm, 10 - 13 Mar. 2002
http://www.isn-oldenburg.de/talks/ulm_maerz02/
- M. Hohlfeld, E. R. Hilf, T. Severiens, H. Stamerjohanns: Verteilte Informationssysteme für die Wissenschaften (und ihre Vernetzung). Vortrag und Paper für den Tagungsband der Konferenz 'Die Zukunft des wissenschaftlichen Publizierens', Forschungszentrum Jülich, 28 - 30 Nov. 2001
http://www.isn-oldenburg.de/~hohlfeld/juelich_nov01.html
- E. Hilf, M. Hohlfeld, T. Severiens, K. Zimmermann: Distributed Information Services in Physics. Veröffentlicht in HEP Libraries Webzine, Issue 4, June 2001 [ISSN 1424-2729]
<http://library.cern.ch/HEPLW/4/papers/2/>
- T. Severiens, M. Hohlfeld, K. Zimmermann, E. R. Hilf: PhysDoc - A Distributed Network of Physics Institutions Documents - Collecting, Indexing, and Searching High Quality Documents by using Harvest, veröffentlicht in D-Lib Magazine, Vol. 6 No. 12, December 2000 [ISSN: 1082-9873]
[DOI: 10.1045/december2000-severiens]
<http://www.dlib.org/dlib/december00/severiens/12severiens.html>
- SINN: Suchmaschinennetzwerk im Internationalen Naturwissenschaftlichen Netz <http://isn-oldenburg.de/projects/SINN/>
PhysNet, Physics Institutions and Documents Worldwide
<http://www.physnet.net>

Open Archives: Distributed services for physicists and graduate students (OAD)
<http://isn-oldenburg.de/projects/OAD>

CARMEN: Content, Analysis, Retrieval and Metadata: Effective Networking
<http://www.mathematik.uni-osnabrueck.de/projects/carmen/index.en.shtml>

W3C XML Query Group
<http://www.w3.org/XML/Query>

Open Archives initiative (OAi)
<http://www.openarchives.org/>

Diesen Text, die zugehörige Präsentation und weitere Informationen finden Sie
auch online unter <http://www.isn-oldenburg.de/~hohlfeld/iuk2003.html> .

Online-Hochschulschriften für die Praxis

Nikola Korb

Die Deutsche Bibliothek

Online Dissertations in Practise

The advantages of electronic publications towards printed documents, i.e. an effective retrieval, are well known. The requirements for exploiting the surplus value seem to be rather unknown, i.e. in structure and semantics. To enable the specific search for quotations, head lines etc., those structural elements need to be appropriately labelled and not just put in italics or bold as it is usual in printed publications. Authoring electronic documents demands more discipline, but the readiness for that is missing. In particular missing is the insight into the necessity for the disciplined use of structural elements.

A possibility to raise the popularity of electronic publications, i.e. online dissertations, is a change of attitude in defining the original document. At present, the top priority is in creating a well readable layout on paper. That priority will last as long as the printed version counts as the original document and the electronic document as a copy. The priorities will change as soon as the electronic version becomes the original, too.

Suggestions for authoring high class electronic publications were being made in the DFG-project „Dissertationen Online“ initiated by the IuK-Initiative and in the „Elektronisches Publizieren an Hochschulen“ –Proposals- by DINI. Those suggestions „only“ need to be put into practice.

The Co-ordination Agency DissOnline has been part of Die Deutsche Bibliothek since February 2001. It is to provide reference in questions concerning online dissertations and post-doctoral theses, to propagandise their spread of use and to coordinate and stimulate the further development of the results of the project „Dissertationen online“.

Mittlerweile werden fast alle Dissertationen am PC erstellt, deshalb bietet es sich an, sie auch online zur Verfügung zu stellen. Die Online-Publikation bietet für Autoren, Wissenschaftler und Bibliothekare einen Mehrwert, z.B. durch:

- die mögliche Integration vielfältiger Darstellungsmöglichkeiten wie 3D- und Farbdarstellungen, Simulationen, Animationen, Video- und Audio-Sequenzen u.a.m.,

- die sofortige Verfügbarkeit des Dokuments oder einzelner Teile davon am Arbeitsplatz des Benutzers und damit die sofortige Überprüfbarkeit von Verweisen und Zitaten,
- die weltweite Vernetzung neuer Forschungsergebnisse zum Thema und die Möglichkeit der gezielten Recherche nach Autor, Titel, Jahr, Schlagworten, Abstract oder im Volltext von Dissertationen auf verteilten Servern,
- die kostengünstige Veröffentlichung,
- die Beschleunigung bibliothekarischer Arbeit durch Datenübernahme und den Wegfall bestimmter Arbeitsgänge sowie eine platzsparende Archivierung.

Die Voraussetzungen für eine effektive Umsetzung der oben genannten Vorteile wurden von der Kultusministerkonferenz geschaffen. In ihrem Beschluss vom 30. Oktober 1997 wurde der Grundsatz festgehalten, dass der Veröffentlichungspflicht von Dissertationen „durch die Ablieferung einer elektronischen Version, deren Datenformat und deren Datenträger mit der Hochschulbibliothek abzustimmen sind“[1] entsprochen werden kann. Seitdem haben zahlreiche Hochschulen diese Möglichkeit in ihren Promotionsordnungen verankert. Sie wird von den Promovenden zunehmend genutzt. Seit Juli 1998 werden an Der Deutschen Bibliothek Online-Dissertationen und -Habilitationen gesammelt, seitdem stieg die durch Hochschulbibliotheken angemeldete Zahl an elektronischen Hochschulschriften stark an[2].

Stichtag	Anzahl der gemeldeten Online-Hochschulschriften
31.12.1998	97
31.12.1999	521
31.12.2000	1.582
31.12.2001	7.392
31.12.2002	13.204
30.04.2003	15.652

Das Projekt „Dissertationen Online“

Damit die neue Form der Online-Veröffentlichung sinnvoll genutzt und ihre Möglichkeiten ausgeschöpft werden können und zur Klärung von organisatorischen Fragen hat die IuK-Initiative die Arbeitsgruppe „Dissertationen Online“ gegründet. Zur Bearbeitung dieses sehr umfangreichen Komplexes hat die DFG das Projekt „Dissertationen Online“ von 1998 bis 2000 gefördert[3]. Zur Um-

setzung der Projektergebnisse und deren Anpassung an zukünftige Entwicklung wurden eine Reihe von Projekt-Empfehlungen erarbeitet:

- Die entwickelten Lösungen sollen allen interessierten Universitäten, Bibliotheken und Verlagen zugänglich gemacht werden.
- Es soll zu einer formellen Beteiligung an Dissertationen Online aufgerufen und dies an bestimmte Grundsätze und Bedingungen geknüpft werden.
- Die Entwicklung technischer Lösungen auf Basis offener, internationaler Standards (Open Archives Initiative etc.) soll weiterverfolgt werden.
- Für Dissertationen entwickelte Lösungen und Verfahren sollen auf elektronische Publikationen generell angewendet werden.
- Es sollen Empfehlungen und praktische Hilfen für das Verfassen, Strukturieren und Dokumentieren von elektronischen Publikationen (Autorenrichtlinien) herausgegeben werden.
- Die Deutsche Bibliothek soll eine Geschäftsstelle einrichten, die die Zusammenarbeit der Kooperationspartner koordiniert.

Der letztgenannten Empfehlung folgend wurde zum 1. Februar 2001 die Koordinierungsstelle DissOnline für die Nutzung und Koordinierung der Weiterentwicklung der Ergebnisse von „Dissertationen Online“ eingerichtet.

Aufgaben und Aktivitäten der Koordinierungsstelle DissOnline

Die Koordinierungsstelle DissOnline sieht ihren besonderen Aufgabenschwerpunkt in der Koordinierung und Beratung der am Verfahren Dissertationen Online beteiligten Institutionen. Innerhalb Der Deutschen Bibliothek ist die Koordinierungsstelle Anlaufstelle für alle Fragen, welche die Ablieferung, Bearbeitung und Archivierung von Online-Hochschulschriften betreffen.

Nach außen nimmt die Koordinierungsstelle DissOnline folgende Aufgaben wahr:

- Sie ist Ansprechpartner und Kontaktstelle für wissenschaftliche Institutionen, Bibliotheken, Verlage und Autoren.
- Sie regt die Kooperation der beteiligten wissenschaftlich publizierenden Einrichtungen untereinander an und fördert den Know-how-Transfer. Sie koordiniert Weiterentwicklungen und Anpassungen, macht sie bekannt und stellt sie zur Verfügung.
- Sie empfiehlt technische Neu- oder Weiterentwicklungen und orientiert sich dabei am „Stand der Technik“ und nationalen sowie internationalen Standardisierungsbestrebungen.

- Die Koordinierungsstelle betreibt Öffentlichkeitsarbeit und fördert das Publizieren von Dissertationen in elektronischer Form. Dazu organisiert die Koordinierungsstelle Informationsveranstaltungen, hält Vorträge und tritt an Messeständen auf.
- Außerdem arbeitet die Koordinierungsstelle mit nationalen und internationalen Organisationen (v.a. der NDLTD) zusammen, um ihre Zielsetzungen zu verwirklichen.

Seit September 2002 ist die Koordinierungsstelle DissOnline auch für Redaktion, Betreuung und Aktualisierung der im DFG-Projekt entwickelten Seiten von <http://www.dissonline.de> - „Digitale Dissertationen im Internet“ verantwortlich. Über diese Informationsbörse erhalten Bibliotheken, Doktoranden, Fachbereiche und Rechenzentren Hinweise zu Tools, Online-Hilfen und Ansprechpartnern sowie die Empfehlungen Der Deutschen Bibliothek zu bevorzugten Formaten und Tools.

Strategien zur Sicherung der Langzeitverfügbarkeit elektronischer Dissertationen

Elektronischen Hochschulschriften sollen nicht nur unter heute bekannten Bedingungen benutzbar sein. Ziel ist es, die elektronischen Hochschulschriften langfristig zu archivieren und die dauerhafte Verfügbarkeit zu gewährleisten. Um die mit der Langzeitarchivierung und –verfügbarkeit verbundenen Probleme zu bewältigen, wurden seitdem in verschiedenen Projekten wie z.B. „Dissertationen Online“, „MathDiss International“[4], „CARMEN“[5] und „META-LIB“[6] Lösungen und Hilfen für Produktion, Retrieval und Archivierung von Hochschulschriften entwickelt. Danach sollten Dissertationen für die Archivierung in ein sicheres Archivierungsformat (z.B. XML), für die Präsentation auf dem Bildschirm in ein Präsentationsformat (z.B. HTML) und für den Ausdruck in ein Layout-orientiertes Druckformat (z.B. PDF) konvertiert werden.

Die Vorteile für die Langzeitarchivierung von strukturierten Dokumentformaten, wie XML gegenüber Layout-orientierten Formaten sind offensichtlich. Die Dateien im XML-Format sind mensch- und maschinenlesbar. XML ist auf verschiedenen Hardwareplattformen und verschiedenen Betriebssystemen verfügbar. Durch einfache Transformationen ist die Erstellung von neuen Präsentationsformaten und Druckformaten möglich.

Multimediaobjekte können integriert werden. Mit SVG und SMIL steht XML auch direkt für Bilder und Filme zu Verfügung. Wenn es notwendig wird, kann durch die Maschinenlesbarkeit und die Auszeichnung der Struktur und nicht des Layouts, vereinfacht automatisch migriert werden. Die automatische Verarbeitung wird durch die Verwendung einer einheitlichen DTD, wie z.B. der

für Dissertationen entwickelten XDiML (<http://edoc.hu-berlin.de/diml/>) unterstützt.

Die DINI-Arbeitsgruppe „Elektronisches Publizieren“ hat Empfehlungen [7] zur Erstellung, Verarbeitung, Sicherheit und Recherche von Dokumenten herausgegeben. Damit wurden die wichtigsten Voraussetzungen für die Langzeitverfügbarkeit von elektronischen Dokumenten zusammengefasst. Besonders sind die Hinweise zur Autorenbetreuung, die benötigten Metadaten und die zu empfehlenden Dokumentformate zu beachten.

Die unten abgebildete Tabelle mit einer Aufschlüsselung der bisher bei Der Deutschen Bibliothek abgelieferten elektronischen Dissertationen zeigt, dass zur Umsetzung der DINI- Empfehlungen (Format XML) noch einiges getan werden muss.

Dateiformat	Anteil in Prozent
PDF	86,1
PS	2,0
Sonstige	11,9

Aufgrund dieser Zahlen sind einige Bemerkung zu Erstellung von PDF notwendig. Nach Möglichkeit sollte strukturiertes PDF (Version 1.4.) erzeugt werden. Dies ist zur Sicherung der Langzeitverfügbarkeit aus oben genannten Gründen besser geeignet als unstrukturiertes PDF.

Durch die Nutzung von Postscript-Standardfonts wird eine zukünftig notwendige Migration vereinfacht. Bei Sonderzeichen, Symbolen usw. reichen die Standardfonts nicht aus. Hier ist die Verwendung von TrueType-Fonts bzw. Type-1-Fonts zu empfehlen. Diese Fonts müssen eingebettet werden, sonst wird auf anderen Rechnerumgebungen automatisch nach Ersatzfonts gesucht. Die Ersatzfonts bilden nicht immer einen identischen Zeichensatz ab, so kann durchaus ein „+“ durch ein „-“ ersetzt werden. Alternativ ohne „geeigneten“ Font auf der neuen Rechnerumgebung bleiben die Seiten leer.

Auf die Verwendung von Type-3-Fonts sollte verzichtet werden. Type-3-Fonts sind Bitmap-Fonts mit allen Vor- und Nachteilen eines Bitmaps. Für die Darstellung eines Dokuments am Bildschirm sind sie ungeeignet. [8]

Ein wesentlicher Vorteil von Online-Dissertationen ist die sofortige Benutzbarkeit. Dazu muss aber die Datei in absehbarer Zeit downloadbar sein. Dazu ein paar Zahlen: 99% aller Dissertationen, die an Die Deutsche Bibliothek geliefert werden, sind kleiner als 50 MB. Es gibt aber auch welche, die größer als 300 MB sind.

Wenn jemand eine Dissertation von 50MB herunterlädt, braucht er mit ISDN ca. 2 Stunden. Eine noch größere Arbeit wird kaum genutzt werden, weil der Download viel zu lange dauert.

Wenn ein Autor möchte, dass seine Arbeit genutzt wird, muss auch auf die Dateigröße geachtet werden.

Deshalb sollten soweit dies möglich ist, nur Vektorgrafiken verwendet werden. Für eine Vektorgrafik wird nur Art, Form und Größe der grafischen Objekte gespeichert. Dabei wird deutlich weniger Speicher als bei Pixelgrafiken, bei denen jeder einzelne Bildpunkt mit Farbinformationen gespeichert werden muss, benötigt. Ist die Einbringung von Pixelgrafiken unumgänglich, reicht womöglich eine geringere Auflösung und weniger Farbtiefe.

Persistent-Identifizierungs-Management (URN)

Zur Nutzung von elektronischen Dokumenten ist es wichtig, dass sie genau wie gedruckte Dissertationen dauerhaft und eindeutig zugriffs- und zitierfähig bleiben.

Gegenwärtig wird die „elektronische Adresse“ von Internet-Dokumenten überwiegend in Form des Uniform Resource Locators (URL) realisiert. Jede URL ist weltweit eindeutig. Aus diesem Grund werden URLs sowohl für den Zugriff auf das Dokument als auch als Identifier für die Zitierung von Publikationen verwendet. Wird jedoch der elektronische Standort des Dokumentes verändert, sind alle Referenzen auf das Dokument nicht mehr benutzbar.

Als Grenzen der URL lassen sich beispielhaft folgende Punkte nennen:

- Methoden der Adressierung, die eine langfristige Lösung versprechen, können sich ändern,
- URLs, die durch einen Linkcheck als fehlerhaft identifiziert wurden, müssten unter hohem Arbeitsaufwand in allen Nachweissystemen wie Katalogen, Bibliografien oder Portalen korrigiert werden,
- Änderungen in gedruckten Publikationen, die Internet-Adressen zitieren, sind i.d.R. unmöglich,
- URLs können temporär durch instabile Verbindungen zu Servern, Netzwerkfehler etc. ausfallen, es ist keine Information verfügbar, ob eine Kopie des Dokumentes existiert, auf die er alternativ zugegriffen werden könnte,
- Elektronische Dokumente verändern auch in verteilten Geschäftsprozessen ihren Standort und können durch URLs nicht zuverlässig identifiziert und adressiert werden.

Eine Lösung ist die Anwendung von Persistent Identifiers. Dadurch sind elektronische Dokumente referenzierbar und auch bei einer sich ändernden URL er-

reichbar. Um einen dauerhaften Zugriff auf elektronische Ressourcen zu gewährleisten, hat Die Deutsche Bibliothek aufbauend auf dem BMBF-Projekt CARMEN AP4[9] seit September 2001 einen Dienst zur Vergabe und Verwaltung von Persistent Identifiers aufgebaut, und verwendet als Persistent Identifier den „Uniform Resource Name“ (URN).

URN's, die an Der Deutschen Bibliothek administriert werden, leiten sich aus dem internationalen Namensraum „NBN“ (National Bibliography Number) ab. Der Namensraum „NBN“ ist eine gemeinsame Initiative der Nationalbibliotheken. Das Konzept sieht eine verteilte Administrierung vor, bei der den einzelnen Nationalbibliotheken die Verwaltung und Auflösung von URN entsprechend ihrem Sammelauftrag zugeordnet wird. Die Deutsche Bibliothek ist damit für die Verwaltung von URN mit dem Namensraum „nbn:de“ verantwortlich.

Eine deutsche URN wird nach dem folgenden Muster gebildet:

urn:nbn:de:[Verbundabkürzung]:[Sigelnummer]-[Nummer]
(Beispiel: **urn:nbn:de:gbv:18-7900**)

Dabei bedeutet [Verbundabk.] ein Buchstabenkürzel des jeweiligen Bibliotheksverbundes, [Sigelnummer] steht für das Bibliothekssigel der Hochschulbibliothek und [Nummer] muss eine eindeutige (Produktions-)Nummer sein.

Gegenwärtig werden im OPAC Der Deutschen Bibliothek keine Herkunfts-URL mehr angezeigt, sondern nur die URN und der Verweis auf die Frontpage des Archivservers Der Deutschen Bibliothek, um die Anzeige ungültiger URLs zu vermeiden und die Vorteile der URN zu nutzen.

Seit September 2001 wurden insgesamt 5.200 URN mit über 30.000 Zugriffen (Stand: 10.05.2003) registriert. Gegenwärtig haben 44 Universitätsbibliotheken ihre Bereitschaft zur Vergabe von URN angemeldet, davon nutzen 30 aktiv dieses Verfahren. Etwa 50 Prozent aller bisher registrierten URN wurden retrospektiv erfasst. Die Tatsache, dass ca. 30 Prozent aller gemeldeten URLs inzwischen – z.B. durch Serverumstellungen bei Universitätsbibliotheken – inaktiv sind, zeigt, dass der Einsatz von URN nötig und sinnvoll ist.

Dieser Persistent-Identifier-Dienst Der Deutschen Bibliothek wird im Rahmen des Projektes EPICUR in eine produktive Systemumgebung überführt. Vergabe und Verwaltung von Persistent Identifiers werden auf weitere Dokumentenklassen ausgedehnt. Nähere Informationen zum URN-Management Der Deutschen Bibliothek sind auf der Homepage des Projektes EPICUR unter <http://www.persistent-identifier.de> verfügbar.

Recherche und Zugriff auf Online-Dissertationen

In einem kurzen Abriss werden verschiedene Möglichkeiten vorgestellt, wie nach Online-Hochschulschriften recherchiert werden kann. Eine Voraussetzung zur Auffindbarkeit einer Dissertation ist bei allen hier beschriebenen Methoden eine Aktivität der Hochschulbibliothek, welche die Online-Hochschulschrift angenommen hat. Diese muss durch Anmeldung des Hochschulschriften-Servers (bei TheO) oder durch Anmeldung jeder einzelnen Dissertation (bei Der Deutschen Bibliothek und in weiteren Archiven) die Dokumente bekannt machen. Angepasst an den Recherchebedarf sollten die unterschiedlichen Zugänge zu den Dokumenten verwendet werden. Der Volltext der Online-Dissertationen ist meist frei zugänglich und kann gemäß dem Urheberrecht zum eigenen wissenschaftlichen Gebrauch verwendet werden.

Aus dem Projekt „Dissertationen Online“ ist die spezielle Dissertationen-Suchmaschine TheO hervorgegangen (<http://www.iwi-iuk.org/dienste/TheO/>).

Diese Suchmaschine greift auf die Metadaten von Online-Dissertationen zu, deren Server bei TheO eingetragen sind. Derzeit werden die Server von 46 Hochschulbibliotheken indiziert. TheO ermöglicht eine komfortable Suche nach Dissertationen durch differenzierte Suchkategorien und -felder. Insbesondere die Recherche nach Dissertationen einzelner Fachgebiete, die zusätzlich noch eingeschränkt werden kann und die Suche im Abstract ist eine der Stärken von TheO.

Im Online-Katalog Der Deutschen Bibliothek (<http://dbf-opac.ddb.de/> oder <http://dbl-opac.ddb.de/>) sind mittlerweile über 15.000 Online-Dissertationen (und über 850.000 gedruckte Dissertationen) verzeichnet

Derzeit haben sich (Stand: 01.04.2003) 78 Hochschulbibliotheken für das Abgabeverfahren von Online-Dissertationen an Die Deutsche Bibliothek angemeldet, von denen 68 aktiv diese Möglichkeit nutzen. Die Suche im DDB-Online-Katalog nach Online-Hochschulschriften ist allerdings etwas kompliziert; insbesondere bei der Einschränkung auf bestimmter Fachgebiete müssen mehrere aufeinander aufbauende Suchschritte erfolgen[10].

Eine weitere Möglichkeit bildet die Suche im OPUS-Informationsverbund (http://elib.uni-stuttgart.de/opus/gemeinsame_suche.php). Die Metasuchmaschine des Stuttgarter OPUS-Projektes indiziert elektronische Dokumente von derzeit 27 OPUS-Anwendern. Gefundenen werden meist nicht nur Dissertationen und Habilitationen, sondern alle Arten von Hochschulschriften wie Reports, Aufsätze und Diplomarbeiten.

Einen globalen Zugang bietet die Suche im „Networked Digital Library of Theses and Dissertations Union Catalog“ (<http://www.ndltd.org/browse.html>). Dieser Gesamtkatalog soll als Depot akademischen Materials dienen; die Beiträge stammen von zahlreichen internationalen Mitgliedsinstitutionen.

Schließlich seien hier noch zwei Beispiele zur Suche bei OAI-Service-Providern erwähnt. Eine Suche über vorwiegend deutsche OAI-Archive bietet der Dokumenten- und Publikationsserver der Humboldt-Universität Berlin (http://edoc.hu-berlin.de/e_suche/oai.php). Dieser indiziert gegenwärtig elektronische Dokumente aus 13 OAI-Archiven und beinhaltet darunter etwa 500 Dissertationen.

Eine weltweite OAI-Suche bietet OAIster, ein Projekt der University of Michigan Digital Library Production Services (<http://oaister.umd.umich.edu/o/oaister/>), welche derzeit über eine Million Datensätze aus rund 150 internationalen Institutionen verzeichnet. Dazu zählen zahlreiche Dissertationen; allerdings kann hier nur im Freitext nach den Begriffen „Thesis“ oder „Dissertation“ gesucht werden.

Statistik

Die große Anzahl der auf dem Archivserver Der Deutschen Bibliothek gespeicherten Online-Hochschulschriften ermöglicht relativ zuverlässige Angaben über deren Aufkommen.

Bei der Recherche nach Online-Hochschulschriften ist ein sehr heterogenes Bild die fachliche Herkunft der Dissertationen betreffend festzustellen. Eine Übersicht[2] verdeutlicht, dass naturwissenschaftliche Fächer einen sehr hohen Prozentanteil der in diesen Fächern angefertigten Dissertationen als Online-Publikation einreichen:

Spitzenreiter ist die Chemie mit fast 45% Online-Dissertationen. Der Durchschnitt über alle Fächer liegt bei ca. 17%. Der Anteil von Online-Dissertationen im geisteswissenschaftlichen Bereich liegt bei ca. 5%.

Aufkommen nach Fach im Jahr 2002 (die obersten 9 Fächer, nach Prozentanteil)	Anzahl der Online-Dissertationen	Online-Anteil in Prozent
Chemie	301	44,96
Veterinärmedizin	133	44,26
Biologie	342	40,52
Mathematik	74	37,37
Physik / Astronomie	342	36,31
Chemische Technik; Lebensmittel-, Textiltechnik	104	25,49
Informatik	67	25,33
Geowissenschaften und Geographie	114	22,53
Medizin	1050	20,94

Allein die Hochschulschriften im Fach Medizin, welche die größte Anzahl der Online-Hochschulschriften bilden, stellen über 30 Prozent aller seit 1998 archivierten Dokumente. Das Fach Informatik, welches in der obigen Übersicht für 2002 den an absoluten Zahlen geringsten Anteil einnimmt, erreicht immer noch fast sechs Prozent. Demgegenüber ist zu konstatieren, dass sämtliche Geistes- und Sozialwissenschaften - zusammen genommen – weniger als zehn Prozent aller archivierten Online-Hochschulschriften bilden.

Ein Schwerpunkt der Koordinierungsstelle ist deshalb die Förderung der Nutzung und Erstellung von Online-Dissertationen gerade auch in den geisteswissenschaftlichen Fächern.

Ausblick

Der kurze Überblick hat gezeigt, dass das Aufgabenspektrum der Koordinierungsstelle DissOnline mit zunehmender Zahl der abgelieferten Dissertationen an Komplexität zunimmt. Um nötige Weiterentwicklungen aufbauend auf den Ergebnisse des Projekts „Dissertationen online“ anzuregen und um ein Kompetenznetzwerk aufzubauen, fördert die DFG seit dem 1. Januar 2003 ein Projekt zum Aus- und „Aufbau einer Koordinierungsstelle für elektronische Hochschulschriften“.

Das Projekt beinhaltet u.a. eine Analyse zur Infrastruktur im Bereich Online-Dissertationen an den deutschen Hochschulen mit Vergleichen von Promotionsordnungen, Geschäftsgängen, Online-Angeboten und –Portalen. Mit den daraus gewonnenen Ergebnissen wird ein interaktives Informationssystem aufgebaut und die Empfehlungen für Promovenden, Fachbereiche, Bibliotheken werden ergänzt und erweitert. Zur Klärung rechtlicher Fragen werden projektbegleitend Werkverträge vergeben.

Ein Ergebnis des Projekts sollen Vorschläge zu einem einheitlichen, praktikablen Verfahren DissOnline in Deutschland stehen, die mit Fachbereichen, Rechenzentren und Bibliotheken diskutiert und abgestimmt werden. Die Ergebnisse und wesentliche Zwischenstände des Projekts – wie etwa die Auswertung der derzeit laufenden Analyse zur Infrastruktur - werden unter <http://www.dissonline.de/umfrage/start.html> veröffentlicht.

Referenzen Stand: April 2003

- [1] <http://www.ub.ruhr-uni-bochum.de/DigiBib/Tauschseiten/KmKBeschluss.pdf>
- [2] http://deposit.ddb.de/netzpub/web_online-hochschulschriften_stat.htm

- [3] http://www.dissonline.de/tagungen/abschlusstagung_2000_12_13/ergebnis.pdf
- [4] <http://www.ub.uni-duisburg.de/mathdiss/>
- [5] <http://www.mathematik.uni-osnabrueck.de/projects/carmen/>
- [6] http://www2.sub.uni-goettingen.de/workshops_de.html
- [7] <http://www.dini.de/documents/DINI-EPUB-Empfehlungen-2002-03-10.pdf>
- [8] Merz, Thomas und Drümmer, Olaf; Die PostScript -& PDF-Bibel, 2.Aufl.; München, Heidelberg 2002; S. 128ff.
- [9] http://www.bis.uni-oldenburg.de/carmen_ap4/index.html
- [10] Eine Anleitung zur Suche nach Online-Dissertationen im OPAC Der Deutschen Bibliothek findet sich unter
http://www.dissonline.de/texte_html/sucheinstiege.html

Infrastrukturen für innovative Digital Library-Dienste

Perspektiven des Kompetenznetzwerks „Neue Dienste, Standardisierung, Metadaten“

Rudi Schmiede, Stephan Körnig

1 Ausgangslage

Entwicklungstendenzen

Die gegenwärtige Entwicklung im Bereich elektronischer Informationssysteme ist von einer kaum noch zu überbietenden Dynamik geprägt. Die immer kürzeren Innovationszyklen im Bereich der Hardware bieten einen technischen Fortschritt, der einen ebenso raschen Fortschritt in der Verbesserung der Nutzbarkeit von Informationssystemen und eine nahezu vollständige Verfügbarkeit aller Arten von Information erwarten lässt. Die gleichzeitige Veränderung der Plattformen, der Produkte und der für ihre Einsatzbereiche definierten Protokolle erzeugen dagegen ein Klima der Investitionsunsicherheit. Die Anforderungen an die Anwendungsentwicklung steigen ebenso - dies nicht zuletzt aufgrund gesteigerter Erwartungen der Nutzer im Hinblick auf die Qualität der angebotenen Dienste (z.B. graphische Nutzerschnittstellen, die von unterschiedlich qualifizierten Anwendern genutzt werden können).

Die zunehmende Vernetzung der Systeme untereinander, auf deren Grundlage die Vision einer globalen Digitalen Bibliothek oder eines „semantischen Netzes“ entstanden ist, ist keineswegs durch die bisher erreichte Standardisierung von Zugriffsverfahren (das sog. „World Wide Web“) abgeschlossen. Sie befindet sich in einer rasanten Entwicklung, die Herausforderungen unterschiedlichster Art mit sich bringt. Gerade im Bereich der wissenschaftlichen Informationssysteme und in den ihm benachbarten Anwendungsbereichen in Schulen, Universitäten oder Unternehmen sind wir noch weit davon entfernt, die zugänglichen Informationen im Hinblick auf inhaltliche Fragestellungen sicher und vollständig nutzen zu können. Die Erschließung, Nutzung, Auswertung und langfristige Absicherung (Langzeit-Archivierung) digitaler Medien stellt nach wie vor ungelöste Probleme, sie wirft neue urheberrechtliche Fragen auf, und

sie schafft einen Markt, der schwer abzuschätzen und mit hohen Risiken behaftet ist.

Strategiestudien

Vor diesem Hintergrund ist es leicht, sich über die „traditionellen“ Formen zu beklagen, in denen sich die gegenwärtige Praxis bewegt, – weitaus schwieriger ist es aber, eine konkrete Strategie zu entwickeln, die einerseits verbindlich genug ist, um Handlungslinien sichtbar werden zu lassen, andererseits aber auch die benötigte Zukunftsoffenheit besitzt.

Mit dem Beginn des Jahres 2001 hat das Bundesministerium für Bildung und Forschung (bmb+f) daher eine Reihe von Studien in Auftrag gegeben, um – neben den bestehenden Erfahrungen – auch auf dieser Basis eine Programmatik erarbeiten zu können. Als größtes Vorhaben wurde die von der Unternehmensberatung Arthur D. Little durchgeführte Strategiestudie „Zukunft der wissenschaftlichen und technischen Information“ vergeben. Weitere Studien befassten sich mit der Nutzung elektronischer Fachinformation an den Hochschulen und mit den Geschäftsmodellen, die im Zusammenspiel von Bibliotheken und Verlagen entwickelt werden können.¹

In diesem Rahmen wurde auch das – von den Verfassern zusammen mit Prof. Wolfram Koch von der GDCh durchgeführte – Projekt „dl-konzepte - bmb+f-Projekt im Digital Library-Forum“ initiiert, das die Aufgabe hatte, Vorschläge für eine zukunftsweisende und angemessene Gestaltung der Fachinformation in den nächsten Jahren zu erarbeiten. Dieses Projekt nahm als Ausgangspunkt seiner Überlegungen die Perspektive der Nutzer von Fachinformation ein. Es wurden die Bedürfnisse bei der Nutzung elektronischer Fachinformation herausgearbeitet und operative Schritte zur Umsetzung der sich hierbei ergebenden Ziele vorgeschlagen. Ein wichtiges Element von dl-konzepte war die Einbeziehung der Fachöffentlichkeit in Form von Workshops und Arbeitsgruppen und damit die Berücksichtigung der hier vorhandenen Expertise. Insbesondere ist hierunter der Arbeitskreis „Infrastrukturen für Digitale Bibliotheken“ zu erwähnen, der durch dl-konzepte und die IuK Initiative eingerichtet und betreut wurde. Vertreter aus unterschiedlichen Einrichtungen und Disziplinen nahmen an diesem Arbeitskreis teil und richteten kleinere Arbeitsgruppen ein, die sich mit spezifischen Themenschwerpunkten auseinander setzten. Ein Ergebnis dieser Aktivität waren eine Expertenrunde zum Thema Langzeitarchivierung Ende 2001, verschiedene Workshops an Der Deutschen Bibliothek und der Aufbau

1 Die Titel dieser Studien lauten: „Nutzung elektronischer Fachinformation, -publikation und -kommunikation in der Hochschulausbildung. Barrieren und Potenziale in der innovativen Mediennutzung im Lernalltag der Hochschulen“ und „Geschäftsmodelle für elektronische Informationsangebote zwischen Verlagen und Bibliotheken“. Berichte und Ergebnisse finden sich auf dem Server www.dl-forum.de.

eines Kompetenznetzwerks zum Thema Langzeitarchivierung, das seit kurzem vom bmb+f gefördert wird.

Nachdem sich im Verlaufe des Projekts abzeichnete, dass fachliche Portale eine zentrale Rolle bei der weiteren Entwicklung spielen werden, wurden zusätzliche Expertisen gezielt eingeholt. Diese Expertisen wurden in der zweiten Hälfte des Jahres 2001 angefertigt und stellen eine „Schnappschuss“ der Entwicklung zu diesem Zeitpunkt dar.²

Diese unterschiedlichen Studien und Aktivitäten stellen, zusammen mit weiteren konzeptuellen Überlegungen, den Hintergrund für das nun vom bmb+f initiierte Kompetenznetzwerk „Neue Dienste, Standardisierung, Metadaten“ dar. Bevor wir auf diese neuere Aktivität zu sprechen kommen, soll ein genauerer Blick auf die „Landschaft“ geworfen werden, in der (und für die) ein solches Kompetenznetzwerk Impulse geben kann und soll. Kompetenznetzwerke dienen ja keineswegs dazu, „Kompetenz“ zu akkumulieren oder so etwas wie den „Hort der Weisheit“ zu institutionalisieren; sie sollen vielmehr in schlanker Organisationsform konkrete Aktivitäten und die an ihnen Beteiligten zusammenbringen und dadurch die Kooperation der aktiven Personen und Teams initiieren und gewährleisten. Es gilt, dafür die sich anbahnenden Entwicklungen zu antizipieren und den gegenwärtig vorhandenen Gestaltungsspielraum zu nutzen.

Diesen Spielraum sehen wir vor allem durch zwei (nicht nur) für die Fachinformation bestimmende Tendenzen geprägt, auf die wir nun ausführlicher eingehen: Die sich auf breiter Front durchsetzenden Web Standards und die zunehmende Akzeptanz „semantischer“ Frage- und Aufgabenstellungen im Informations- und im Wissensmanagement.

Web Standards

Die USA schienen mit ihren ehrgeizigen Digital Library Projekten seit Mitte der neunziger Jahre Europa schon um einige Jahre voraus zu sein. Dieser Eindruck hat sich mittlerweile relativiert. Dies liegt daran, dass mit der Verbreitung von Webtechnologien und der Erhöhung der Übertragungsgeschwindigkeiten durch die Bandbreite im Internet die aufwendig entwickelten technischen Infrastrukturen und Protokolle für den Alltag des Wissenschaftlers obsolet geworden sind (ausgenommen natürlich die Bereiche, in denen überaus hohe Bandbreiten nach wie vor notwendig sind, z.B..im Grid Computing). Die Nutzungsplattform der Wissenschaft und die des durchschnittlichen „Surfers“ unterscheidet sich in der Regel nicht mehr.

Durch dieses Zusammenrücken der technischen Horizonte agieren die Anbieter elektronischer Fachinformation und die Förderer in einem Markt, der für die unterschiedlichsten Anbieter von Interesse ist. Die bedeutenden Standardi-

2 Die Ergebnis-papiere von dl-konzepte können eingesehen werden unter http://www.dl-forum.de/Foren/dl_konzepte/index.asp

sierungsprozesse werden deshalb auch durch Aktivitäten größerer Unternehmen mit vorangetrieben. Dies betrifft insbesondere die Web-Standards (W3C) und die Protokolle (IETF). Speziellere Standards, die für diesen Markt nicht eine zentrale Bedeutung besitzen, werden dagegen stärker von den sie nutzenden „communities“ geprägt. Im Bereich der Bibliotheken hat sich Dublin Core als Standard für die Metadaten „im Web“ etabliert, während die Verlage auf DOI setzen, das neben einer (bisher) einfach gehaltenen Erschließung die marktrelevanten Fragen der Verfügbarkeit und der Zuordnung von Content zu den Rechteinhabern zu verwalten hilft. Die Bedeutung von Z39.50 als ein Protokoll zum Zugriff auf bibliographische Daten besteht weiterhin - sie wird aber in dem Maße zurückgehen, in dem entsprechende W3C-Standards entwickelt werden, um den Informationsaustausch zwischen verteilten Applikationen zu gewährleisten.

Bedingt durch diese Situation bestehen für europäische Projekte hervorragende Chancen, die weitere Entwicklung wissenschaftsrelevanter Fachinformationssysteme mitzugestalten. Hier ist die besondere Expertise im Bereich der Erschließung, die traditionell intensivere Beschäftigung mit den Fragen der Semantik und die Erfahrungen mit multilingualen Systemen durchaus ein gewisser Vorteil gegenüber den bisher stärker technisch geprägten Ansätzen in den USA. Dieser Vorteil muss allerdings auch - etwa durch eine Präsenz in den relevanten internationalen Arbeitsgruppen - eingespielt werden.

Semantic Web

Das W3C-Konsortium, das mit der Normierung und Weiterentwicklung von Webstandards befasst ist, hat seit gut zwei Jahren die „Semantic Web“ Aktivität verstärkt (die Leitung dieser Aktivität übernahm Eric Miller, der zuvor intensiv an der Weiterentwicklung des Dublin Core Standards beteiligt war). Hierdurch wird es wahrscheinlich, dass die im Bibliotheksbereich bedeutende Dublin Core Initiative durch W3C Standards unterstützt wird. Dies wird vermutlich dazu führen, dass der Zugriff auch auf komplexere Metadatenstrukturen durch international standardisierte Verfahren verbessert wird.

Flankiert wird diese Maßnahme seit Beginn des Jahres 2002 durch die „Web Services“-Aktivität. Hier wird an der Interoperabilität im Hinblick auf die Inanspruchnahme von Diensten durch standardisierte Protokolle gearbeitet. Im Zentrum des Interesses steht das SOAP-Protokoll, das als Middleware die Verarbeitung von Informationen in verteilten Systemen ermöglicht. So könnte - um die Möglichkeiten dieser neuen Architektur an einem Beispiel zu illustrieren - bei der Recherche in einem Archiv (d.h. bei einem Informationsanbieter) der Thesaurus eines anderen Anbieters (eines Diensteanbieters) benutzt werden, um die Suche in diesem Archiv zu verfeinern.

Beide Aktivitäten sind, auch wenn sie an die Probleme der maschinellen Verarbeitung von „Bedeutungen“ rühren, nicht als Grundlagenforschung anzusehen. Vielmehr geht es darum, im Web die Verknüpfung von semantisch reichhaltigen Informationen zu unterstützen, ohne durch eine Strategie des „kleinsten gemeinsamen Nenners“ die Mehrwerte, die spezialisierte Services bieten, schon vom Ansatz her zu verspielen.

2 Perspektiven einer Strategie zur nachhaltigen Gestaltung

Strategische Ansätze

Die Vorschläge, die im Rahmen von dl-konzepte erarbeitet wurden, stützen sich auf diese deutlich wahrnehmbaren Entwicklungstendenzen, die geeignet sind, die künftige Form elektronischer Fachinformation nachhaltig zu verändern und ihre Leistungsfähigkeit erheblich zu verbessern. Es sind *Standardisierungsprozesse* in Gang gekommen, auf deren Grundlage sich XML-basierte Formate zur Strukturierung und für den Austausch von Informationen auf breiter Front durchsetzen. Zweitens gewinnt „Meta-Information“ - vor allem durch die Zunahme der Kommunikation von Maschine-zu-Maschine - zunehmend an Bedeutung. Um zu betonen, dass diese Entwicklung mit dem klassischen Begriff „Metadaten“ kaum noch angemessen zu beschreiben ist, ziehen wir die Bezeichnung *Informationen über Informationen* vor. Drittens wird das sog. *Semantic Web* zunehmend als Entwicklungsperspektive für innovative Formen der Gestaltung von Informationsangeboten angesehen. Entsprechend dieser veränderten Ausgangslage werden sowohl an die Gestaltung von Informationsangeboten als auch an die Funktion wissenschaftlicher Portale neue Anforderungen gestellt.

Ein strategischer Ansatz unter der Berücksichtigung dieser Entwicklung muss demnach darauf abzielen, Projektarbeit zu initiieren, zu intensivieren oder zu unterstützen, die diese sich etablierenden Standards in realen Anwendungskontexten implementiert. Dies ist insbesondere deshalb vielversprechend, weil diese Standardisierung (insbesondere im Rahmen des W3C) auf Interoperabilität der verschiedenen Aspekte der Informationsmodellierung zielt. Zudem unterstützen die XML-Formate die (semantisch reichhaltige) mittel- und längerfristige Sicherung von Daten, Dokumenten und Metadaten. Investitionen in die Umstellung auf diesen Standard werden also nicht nur im Kontext einzelner Projekte sinnvoll sein; es zeichnet sich vielmehr ab, dass auf diese Weise auch die längerfristige Nutzung im Rahmen bisher noch nicht spezifizierter Szenarien sichergestellt werden kann.

Handlungsfelder

Es wurden drei zentrale Handlungsfelder identifiziert, die eine besondere Hebelwirkung im Sinne dieser strategischen Ausrichtung versprechen. Erstens werden *Standardisierungsprozesse* als ein Kernelement der wissenschaftlichen und technischen Innovation angesehen. Zweitens besteht ein großer Bedarf an der *Spezifikation, der Standardisierung und der Implementierung digitaler Dienste*, die neue Möglichkeiten für die Schaffung *fachlicher Zugänge* (fachlicher Portale) bieten. Drittens sollten im Bereich der *lokalen und regionalen Zugänge* Verbesserungen erzielt werden, die insbesondere auf die folgenden Aspekte abzielen: Die Unterstützung dezentraler Organisationsformen, die Integration vorhandener Angebote und die Verbesserung der „Zugangsseite“ einschließlich der Authentifizierungsaspekte. Diese Handlungsfelder wurden gemeinsam mit Vertretern unterschiedlicher Institutionen diskutiert, und es konnten entsprechende Aktionslinien formuliert werden. Sie sollen im Folgenden skizziert werden.

Handlungsfeld Standardisierung: „Standardisierung ist Forschung“

Die Bedeutung von Standardisierungsprozessen für die weitere Entwicklung digitaler Informationsangebote kann aus Sicht der Experten gar nicht hoch genug eingeschätzt werden. Denn die Etablierung von Standards ist die Grundbedingung für die Realisierung offener und freier Zugangsmöglichkeiten. Entsprechend wurde schon in der ersten Sitzung des AK „Infrastrukturen“ hervorgehoben, dass Standardisierung als integraler Bestandteil der Forschung zu begreifen und zu fördern ist. Im Ergebnispapier des Workshops „Standardisierungsaktivitäten in Deutschland“ vom 10.11.2000 heißt es deswegen:

„Das in Europa vorherrschende Bewusstsein, die Normung sei gegenüber der Forschung ein nachgeordneter Prozess, an dem man sich nicht unbedingt beteiligen müsse, hat zu der beklagenswerten Situation geführt, dass Standards vornehmlich in den USA definiert werden. Europa ‚schreibt dann Jahre später ab‘ und muss sich nach einem Markt richten, der sich mitunter in eine völlig andere Richtung weiterentwickelt hat. Abgesehen davon, dass auf diese Weise kein prägender Einfluss auf technologische Entwicklungen möglich ist, muss festgestellt werden, dass hierdurch erhebliche Eigeninvestitionen verloren gehen. Standardisierungsarbeit muss deshalb als Forschungsarbeit erkannt und anerkannt werden: Es muss eine Bereitschaft bestehen, kurzfristig erhebliche Ressourcen in die Standardisierungsarbeit zu stecken, und es müssen die Personen (möglichst kontinuierlich) an dieser Arbeit beteiligt sein, die in der Forschungsarbeit involviert sind. Beides stellt unabdingbare Voraussetzungen dafür dar, dass die Ergebnisse der Forschung in Europa in

marktfähige Produkte umgesetzt werden können. Fördermittel für diesen Bereich sollten möglichst kurzfristig und unbürokratisch zur Verfügung gestellt werden können.“

Es wurde über die Erfahrungen mit und in internationalen Gremien berichtet: Europäische Vertreter werden oft gar nicht entsandt, da sich erst für sehr konkrete Gespräche, von denen unmittelbar Ergebnisse erwartet werden, Finanzierungsmöglichkeiten ergeben. Hier wird aus unserer Sicht am falschen Ende gespart. Das „Return on Investment“ sollte allen Partnern deutlich sein - selbst wenn die Ergebnisse noch nicht klar abzusehen sind - und eine unterstützende Infrastruktur geschaffen werden. Es müssen die Personen am Standardisierungsprozess teilnehmen, die innovativ tätig sind, richtungsweisende technische Bausteine entwickeln und über die Ressourcen des Forschungsbereichs entscheiden.

Auf dieses Handlungsfeld bezogene Aktionslinien sollten deswegen beinhalten, bzw. es sollten entsprechend gefördert werden:

- die Mitarbeit und Unterstützung der Mitarbeit deutscher und europäischer Arbeitsgruppen an der Formulierung und Verbreitung internationaler Standards
- die Unterstützung von Projekten, die ihre Ergebnisse in Standardisierungs-Gremien einbringen (Setzen von Standards durch aktive Entwicklungsarbeit)
- die kooperative Auswahl wichtiger Services und Standardisierungsbereiche
- die Nutzung von Standards in inländischen Projekten und Entwicklungen, insbesondere die einheitliche Nutzung durch die verschiedenen Anwendergruppen (Wissenschaften, Universitäten, Fachinformationseinrichtungen, Bibliotheken, Verlage, Wirtschaft etc.)
- die Schaffung eines Kompetenz-Netzwerkes mit einer Mitgliedschaft beim W3C und in anderen einschlägigen Initiativen
- der Aufbau eines Informationspools und einer Kooperationsplattform, z.B. auf dem DL-Forum-Server (www.dl-forum.de)

Handlungsfeld fachliche Zugänge zur elektronischen Fachinformation – Portale

Im Hinblick auf die (elektronische) Fachinformation unterscheiden sich in den unterschiedlichen Fächern die Bedürfnisse, der Entwicklungsstand wie auch die Bereitschaft zur Initiative ganz erheblich; es war und ist ja eines der grundlegenden Ziele der IuK-Initiative, durch die Förderung der fachübergreifenden Kooperation diese Differenzen abzumildern. Einheitslösungen in diesem Bereich werden selbst durch noch so gut gelingende Standardisierungsprozesse nicht entwickelt werden können. In diesem Zusammenhang muss noch einmal die Stossrichtung des Konzepts der W3C-Initiative (semantic web) hervorgehoben

werden: Die Idee besteht nicht darin, eine perfekte technische Plattform - das „alles wissende“ Web - zu implementieren; es geht vielmehr darum, für die unterschiedlichen Angebotsformen technisch angemessene Instrumente zur Strukturierung und zum Austausch von Informationen zu entwickeln.

Ebenso ist es zutreffend, dass voneinander unabhängige Entwicklungen Doppelarbeiten und „Atavismen“ (Nutzung veralteter Technologien trotz bereits bestehender Lösungen in anderen Fächern) mit sich bringen. Dem sollte durch die Unterstützung entsprechender Kommunikation (Forum der Fachöffentlichkeit) sowie durch Konzepte für die Nachnutzung bereits zur Verfügung stehender Lösungen und die Entwicklung generischer Modelle entgegengewirkt werden. Unter einem generisches Modell werden hierbei die Bereitstellung von allgemein benötigten Funktionalitäten - eines „Frameworks“, das über Parameter den jeweiligen Bedürfnissen angepasst werden kann, sowie von Schnittstellen zur Einbindung weiterer Funktionalitäten - verstanden. Ebenso sollten insbesondere die Anstrengungen in der EU, die auf die Portalentwicklung gerichtet sind, beachtet und die dort entwickelten Konzepte bei nationalen Projekten berücksichtigt werden.

Mit den Informationsverbünden des bmb+f wird dem Bedürfnis der Fächer nach Bereitstellung von Fachinformation unter den für diese Fächer jeweils geltenden Bedingungen in ersten wichtigen Schritten Rechnung getragen. (U.a. werden hier elektronische Zeitschriftenartikel der jeweiligen Fachverlage verfügbar und/oder durch die von den beteiligten Einrichtungen erzeugten Metadaten recherchierbar gemacht.). Ebenso unterstützt der Zusammenschluss der Verbünde in einem zentralen Verbundgremium (vascoda) hoffentlich Prozesse der Standardisierung und der Qualitätssicherung. Aus unserer Sicht kann sich die Aufgabenstellung für fachliche Portale allerdings nicht darin erschöpfen, in der Hauptsache Zeitschriftenartikel bereitzustellen. Benötigt wird eine Integration aller für das Fach relevanten Angebote - so weit dies unter den gegenwärtigen Bedingungen möglich ist. Es gilt, sowohl die Möglichkeit zur Integration von Verlagsangeboten in Wissenschaftliche Portale als auch Anreize zur Entwicklung innovativer Konzepte zu schaffen. Die Voraussetzung für den Aufbau entsprechender Portale besteht - auch nach Überzeugung der Verlage - in einer genauen Kenntnis der fachlichen Bedürfnisse. Wir schlagen deshalb die Bildung fachwissenschaftlich besetzter Kommissionen vor, die fachliche Anforderungsprofile erarbeiten sollten. Zugleich sollten aber auch konkrete innovative (und generische) Konzepte, so weit diese sich in den einzelnen Fächern abzeichnen, umgesetzt und fachlich evaluiert werden.

Auch die jüngste Aktivität des W3C-Konsortiums im Bereich *Web Services* dokumentiert die Bedeutung von webbasierten Diensten für die weitere Entwicklung von Informationssystemen. Aus unserer Sicht bietet eine möglichst breite Kooperation auf nationaler Ebene die besten Chancen, auf die Entwicklung im internationalen Kontext wirken zu können.

Wir plädieren für die Schaffung eines „offenen Marktes“ für innovative DL-Dienste durch die Etablierung von Schnittstellen. Hierbei geht es nicht um die Entwicklung von neuen Architekturen, sondern um die Realisierung einer modularen und offenen Struktur, die auf Basis von international anerkannten W3C-Standards zu etablieren ist.

Ausgehend von den Aktivitäten des (von dl-konzepte und der IuK-Initiative gemeinsam ausgerichteten) Arbeitskreises „Infrastrukturen für Digitale Bibliotheken“, gilt es nun,

- die jeweiligen Dienste durch Berücksichtigung der (je nach Fach und nach Institution) unterschiedlichen Nutzungsinteressen zu spezifizieren,
- gemeinsam mit den beteiligten Einrichtungen die Nutzung bestehender Standards für die Implementierung entsprechender Dienste zu vereinbaren,
- Konzepte für Bereiche zu entwickeln, in denen entsprechende Standards noch nicht implementiert sind, und
- die Integration dieser unterschiedlichen Spezifikationen durch ein generisches Modell für fachliche Portale zu gewährleisten, das von den Beteiligten gemeinsam getragen wird.

Weitere strategische Ansatzpunkte bestehen darin,

- bei der Weiterentwicklung des Protokolls der Open Archives Initiative mitzuwirken und dieses durch konkrete Projektarbeit im Hinblick auf die Berücksichtigung von DL-Diensten mitzugestalten,
- gezielt Projekte anzustoßen, die in realen Anwendungskontexten Portale mit Schnittstellen für innovative Diensten aufbauen, und
- als eine flankierende Maßnahme ein Kompetenz-Netzwerk zu schaffen, das bestehende und zielführende Aktivitäten moderiert, diese Aktivitäten nach außen sichtbar macht und zur Bewusstseinsbildung in der nationalen wie der internationalen Fachöffentlichkeit () beiträgt.

Handlungsfeld lokale und regionale Zugänge

An den Universitäten ist eine Vielfalt, zugleich aber auch ein Wildwuchs forschungsrelevanter Informationsangebote festzustellen - zahlreiche, aus den konkreten Forschungstätigkeiten und durch Drittmittel generierte Informationsangebote werden entwickelt, ohne dass diese im Sinne eines integrierten Informationsangebotes der Universität erschlossen, zusammengeführt und gepflegt werden. Dies gilt ebenso für regionale Angebote, die mitunter nur wenigen Personen bekannt und oftmals nach außen wenig sichtbar sind. Es gilt, die lokalen Angebote in eine Struktur bringen, d.h. das jeweilige Profil der Institution oder Organisationseinheit herauszuarbeiten und nach außen sichtbar zu machen.

Ziel der gemeinsamen Anstrengung muss es sein, eine Informationsversorgung an den Hochschulen aufzubauen, die einerseits dezentral organisiert ist,

andererseits Kompatibilität und Interoperabilität durch gemeinsame Entwicklung und den breiten Einsatz von Standards sichert. Als strategische Aufgabe muss die Verantwortung für die lokale Infrastruktur zudem in den Hochschulleitungen hochrangig angesiedelt und durch Informationsbeauftragte der Fachbereiche unterstützt werden.

Die verschiedenen Einrichtungen an den Hochschulen sollen – in Kooperation mit weiteren Anbietern bzw. Interessierten an elektronischer wissenschaftlicher Fachinformation – gemeinsame Dienstleistungen aufbauen. Bei der technischen Umsetzung, die möglichst durch verfügbare und erprobte Software erfolgen sollte, sind überregionale und internationale Standards (z.B. XML, RDF, OAP usw.) zu berücksichtigen, um eine Interoperabilität zu gewährleisten.

Den lokalen Zugängen an den Hochschulen oder anderen Einrichtungen in Forschung und Lehre kommt schließlich eine bedeutsame Funktion in der Bündelung der Zugänge und Berechtigungen zu. Einem Nutzer ist es kaum zuzumuten, mit Dutzenden unterschiedlicher Zugänge und Berechtigungen sowie unterschiedlichen Authentifizierungsmechanismen im Alltag umzugehen. Da die Rechte in vielen Fällen institutionell vermittelt sind, sollten mit den institutionsgebundenen Zugängen auch ein einheitliches, möglichst einfaches und nutzerorientiertes Zugangs- und Rechtemanagement verbunden sein.

An dieser Stelle sollte – zur Vermeidung von Missverständnissen – betont werden, dass der von uns charakterisierte strategische Ansatz sich nicht als eine „Neuerfindung“ der Informationslandschaft versteht. Es geht hierbei darum, sich auf die veränderte Problemlage einzustellen und die Prozesse zu unterstützen, die aus der Perspektive der Nutzer wissenschaftlicher Fachinformation wünschenswert sind. Insbesondere muss hierbei gesehen werden, dass sich eine technische Infrastruktur nicht „von oben“ durchsetzen lässt. Die Schaffung des Kompetenznetzwerks sollte als ein weiterführender, möglicherweise vielversprechender Ansatz gesehen werden, dieser Problemlage Rechnung zu tragen.

3 Kompetenznetzwerk „Neue Dienste, Standardisierung, Metadaten“

Konzept

Unter einem Kompetenznetzwerk wird, wie oben schon angedeutet, eine arbeitsteilige, schlanke Form von Organisation verstanden. Die Arbeitsgegenstände des Kompetenznetzwerks werden durch Projektgruppen angegangen, die sich (über einen begrenzten Zeitraum) konkreten Gestaltungsaufgaben zuwenden. Hierdurch gewinnt die gewählte Organisationsform die Beweglichkeit, ihre Ziele in dem Umfeld eines durch beständige Innovation gekennzeichneten Marktes verfolgen zu können. Das Kompetenznetzwerk ist in dieser Hinsicht

auch als ein „Wettbewerb guter Ideen“ zu verstehen, in dem die Kontinuität nicht durch die Beteiligten, sondern durch die Zielsetzungen vorgegeben ist. Dieser durch Inhalte vorgegebene Rahmen wird ergänzt durch Aktivitäten, die auf die Schaffung und Erhaltung eines Forums der Fachöffentlichkeit zielen (Öffentlichkeitsarbeit, Kongresse, Einrichtung von Arbeitsgruppen usw.).

Die Kriterien für geeignete Projektaktivitäten sind nicht festgeschrieben. Beispielfhaft genannt werden können

- die Ausrichtung auf verbesserte und innovative Nutzungsmöglichkeiten der unterschiedlichsten Formen elektronischer Fachinformation (neue Dienste),
- die Orientierung an internationalen Standards (bzw. die Beteiligung an internationalen Standardisierungsprozessen),
- die Verbesserung der Interoperabilität von Informationssystemen über den Aufbau von (möglichst reichhaltigen) „Informationen über Informationen“ (Metadaten, Klassifikationen, Ontologien usw.) und
- die Einbettung der Projektarbeit in reale Anwendungskontexte.

Aktuelle Projekte

Zur Zeit werden im Rahmen des Kompetenznetzwerks vom bmb+f vier Projekte gefördert. Diese Projekte bieten ein breit gefächertes Spektrum an Aktivitäten, auf deren Grundlage innovative Dienste geschaffen werden können.

Eines dieser Projekte ist selbst ein Kompetenznetzwerk. Der Bereich *Langzeitarchivierung* bedarf im besonderen Maße der Etablierung von Kommunikation und einer abgestimmten Rollenverteilung. Die Deutsche Bibliothek, die Bayerische Staatsbibliothek, die Humboldt-Universität Berlin und die Niedersächsische Staats- und Universitätsbibliothek Göttingen bauen zusammen mit dem Bayerischen Staatsarchiv und dem Institut für Museumskunde eine Kommunikationsplattform auf. Diese soll einerseits zu einer dauerhaften und sich selbst tragenden Infrastruktur führen, andererseits sollen auf der Grundlage einer Analyse der Ausgangssituation Projekte angestoßen werden.

Das Konrad Zuse Zentrum für Informationstechnik in Berlin befasst sich in seinem Projektbeitrag mit den Problemen der Suche, des Zugriffs und der Abrechnung in einem größeren Verbund von Anbietern. Es wird ein *dezentraler Zeitschriftenserver* aufgebaut, der innovative Dienste bietet, insbesondere ein auf Open Linking basierendes Referenzsystem und die Integration des „Invisible Web“ sowie weiterer frei zugänglicher Dokumente vorsieht. Die hier gebotenen Möglichkeiten der Suche, der Navigation und des Zugriffs werden insbesondere im Hinblick auf die Bedürfnisse des Wissenschaftlers modelliert werden.

Im Projekt *EPICUR* (Enhancement of Persistent Identifier Services - Comprehensive Method for Unequivocal Resource Identification) der Deutschen Bibliothek wird an einem wichtigen Element der technischen Infrastruktur elek-

tronischer Fachinformation gearbeitet: Der Möglichkeit der dauerhaften Identifikation von Ressourcen im Internet. Das Problem „flüchtigen“ Internetadressen beeinträchtigt zunehmend die wissenschaftliche Kommunikation - ein zuverlässigerer Referenzmechanismus wird benötigt. Im Anschluss an ein Teilprojekt im Verbundprojekt CARMEN im Rahmen von Global Info wird die technische und organisatorische Infrastruktur zum URN-Management weiter ausgebaut, und es wird an der Integration dieses Systems mit anderen Verfahren (insbesondere DOI) gearbeitet.

Das vierte Projekt des Kompetenznetzwerks zum Thema „*Generische und komponentenbasierte Wissenschaftliche Portale* - ein Framework für den Aufbau nutzerorientierter Dienste“ (TU Darmstadt) befasst sich mit einem weiteren Aspekt innovativer Fachinformation - dem standardisierten Austausch strukturierter Informationen auf der Basis des Web Services Standards (W3C). Informationsangebote, die als Dienste spezifiziert und über sog. Registries recherchierbar sind, können auf vielfältige Art und Weise integriert und neuen Nutzungsmöglichkeiten zugeführt werden. Sowohl der Aufbau von Portalen (die auch gegenseitig auf sich als Diensteanbieter zurückgreifen können) als auch der Aufbau von Wissenschaftsnetzwerken soll durch das Framework auf der Grundlage eines sich zunehmend verbreitenden Standards vereinfacht und unterstützt werden. Darüber hinaus hat die TU Darmstadt eine moderierende Funktion im Rahmen des Kompetenznetzwerks übernommen und wird sich um eine Fortführung der in dl-konzepte aufgebauten Kommunikationsplattform für die Fachöffentlichkeit bemühen.

Perspektiven

Die vier genannten Projekte befassen sich mit unterschiedlichen Aspekten einer sich entwickelnden Infrastruktur für fachwissenschaftliche Informationen. Sie haben sich hierbei in besonderer Weise auch mit den internationalen Entwicklungen auseinanderzusetzen, die entscheidend - vor allem durch die Standardisierung im W3C und durch weitere Gremien - die Lebensfähigkeit eines Ansatzes bestimmen. Eine Mitwirkung an diesen internationalen Prozessen wird darüber hinaus auch maßgeblich dafür sein, welche Technologien (und damit welche Investitionen) sich bewähren. Darüber hinaus ist es wichtig, die eigenen fachlichen Bedürfnisse in diesen Prozess mit einzubringen, damit aus der im Prinzip nützlichen und unverzichtbaren Schaffung international geteilter Standards kein Hemmnis für die nutzerorientierte Weiterentwicklung digitaler Dienste entsteht.

Im Herbst 2003 ist deshalb unter dem Titel „DL-Konzepte“ ein Workshop geplant, der zur Beteiligung an der Konzeptualisierung Digitaler Dienste aufrufen, aktuelle Problemstellungen diskutieren und zur Teilnahme an dem sich im Aufbau befindlichen Kompetenznetzwerk motivieren soll.

Das Psychologische Datenarchiv PsychData

Zur Frage der angemessenen Dokumentation von Primärdaten der empirisch-psychologischen Forschung

Jutta von Maurice

Zentrum für Psychologische Information und Dokumentation (ZPID), Trier

Ziel des Psychologischen Datenarchivs PsychData ist die umfassende Dokumentation und langfristige Archivierung von Primärdaten aus allen Bereichen der psychologischen Forschung sowie die Bereitstellung dieser Datensätze für wissenschaftliche Nutzungszwecke. Der vorliegende Beitrag setzt sich näher mit dem grundlegenden Problem der Dokumentation der Archivierungsobjekte auseinander.

1 Daten im empirisch-psychologischen Erkenntnisprozess

Daten bilden die Grundlage empirisch-psychologischer Erkenntnis, die auf das Erleben und Verhalten des Menschen ausgerichtet ist. Obgleich von ihrem Wortursprung auf das Gegebene verweisend, sind Daten erst in einem fortgeschrittenen Stadium des forschungsbasierten Erkenntnisprozesses anzusiedeln, der sich – zumindest idealtypisch – in klar voneinander abgrenzbare Schritte untergliedern lässt (vgl. auch Selg, Klapprott & Kamenz, 1992): Ausgangspunkt jeder empirisch-psychologischen Forschung ist die Umgrenzung der Untersuchungsfragestellung, welche die relevanten Konzepte, die avisierte Population und die postulierten Relationen möglichst präzise umschreibt. Vor diesem theoretischen Hintergrund kann ein zur Bearbeitung der Untersuchungsfragestellung taugliches Erhebungsinstrument ausgewählt oder entwickelt werden. In der empirisch-psychologischen Forschung nehmen Befragungsinstrumente, testdiagnostische Verfahren und experimentelle Designs eine besondere Position ein. Aus der Anwendung dieser Instrumente auf eine Stichprobe resultieren Daten zum Erleben und Verhalten der berücksichtigten Probanden. Die statistische Analyse – als auf diesen Daten aufbauender Forschungsschritt – führt schließlich zur Ableitung des Forschungsergebnisses, das nun nähere Erkenntnis über die eingangs umschriebene Untersuchungsfragestellung verspricht.

Obgleich der Datenbegriff in der Psychologie damit wesentlich umgrenzter genutzt wird als etwa im bibliographischen Kontext, zerfällt er auch hier in mindestens drei Aspekte: Die Konfrontation der Probanden mit dem Erhebungsin-

strument liefert zunächst die so genannten Ursprungsaufzeichnungen oder Rohdaten etwa in Form ausgefüllter Fragebogen, Zeichnungen, Testbögen oder auch in Form von Mausklicks bei computergestützten Erhebungen. Zur Vereinfachung der Auswertung werden diese Ursprungsaufzeichnungen nach festgeschriebenen Regeln kodiert und in eine maschinenlesbare Form übertragen. Häufig findet eine Abbildung der Ursprungsaufzeichnungen in den natürlichen oder reellen Zahlenraum statt, wobei messtheoretische Fragen in den Mittelpunkt rücken. Diese Primärdaten liegen im Forschungsalltag zumeist in Form einer oder mehrerer Datenmatrizen vor, die von den Untersuchungseinheiten und den Untersuchungsvariablen aufgespannt werden. Im weiteren Forschungsprozess werden diese Daten in der Regel mehrfachen Variablentransformationen oder Variablenaggregationen unterzogen (z. B. Umpolungen, Zusammenfassung von Antwortalternativen, Bildung von Skalen und Indizes). Diese abgeleiteten Daten sind mit den zugrundeliegenden theoretischen Konzepten oder mit spezifischen Analyseverfahren häufig eng verwoben.

Insgesamt erweist sich damit der Datenbegriff als in den theoriebasierten, ergebnisorientierten Forschungsprozess komplex eingebunden und – mit der Unterscheidung von Ursprungsaufzeichnungen, Primärdaten und abgeleiteten Daten – als seinerseits mehrdeutig.

2 Theoretische Bewertung und praktische Handhabung von Forschungsdaten

In theoretischer Hinsicht ist die zentrale Rolle von Daten im empirisch-psychologischen Erkenntnisprozess unbestritten. Die moderne Psychologie ist in ihren Grundzügen eine empirische Wissenschaft. In diesem Zusammenhang bleibt – neben der beeindruckenden Quantität empirisch-psychologischer Forschungsdaten – auf die hohe Qualität des Datenmaterials hinzuweisen, die nicht zuletzt auf die theoretische Fundierung des Datenerhebungsprozesses in der Psychologie zurückreicht. Entsprechend konstatiert Erdfelder (1994): „In dem Maße, in dem sich die Psychologie als Erfahrungswissenschaft etabliert hat, ist die Erzeugung und Verwendung empirischer Daten für Lehrende, Studierende und Anwender des Faches zu einer Selbstverständlichkeit geworden“ (S. 47).

Umso bedauerlicher ist, dass dieser aus theoretischer Perspektive hohen Wertschätzung empirischen Materials die Erfahrung entgegensteht, dass Archivierung und Pflege der Datenbasis in der Praxis nicht immer mit der gebotenen Sorgfalt betrieben wird. Gerade die Einbindung in den Forschungsprozess hat zur Folge, dass Primärdaten häufig nicht dauerhaft abgelegt oder zumindest mit abgeleiteten Daten vermischt werden. Darüber hinaus gehen die zumeist im auswertungsrelevanten Format abgelegten Forschungsdaten durch System- und Formatwechsel, Unachtsamkeit und Fremdverschulden viel zu oft innerhalb

kurzer Zeit verloren. All diese Mängel mögen darin begründet liegen, dass nicht nur in der Psychologie, sondern auch in anderen empirischen Forschungsbereichen das Augenmerk häufig einseitig auf die Forschungsergebnisse und deren Publikation gelegt wird und eine Beschäftigung mit der Datenbasis demgegenüber als wenig lohnend erscheint.

In der psychologischen Forschungsgemeinschaft entwickelt sich erst in den letzten Jahren das Bewusstsein, Forschungsdaten selbst als Gut des wissenschaftlichen Austausches anzusehen, dessen Verlust mit hohen Kosten verbunden ist und mitunter der eigenen Glaubwürdigkeit schadet. Entsprechend gelten die von der Deutschen Forschungsgemeinschaft (DFG) 1998 ausgesprochenen Vorschläge zur Sicherung guter wissenschaftlicher Praxis inzwischen auch in der Psychologie als allgemein akzeptierte Leitlinie. Dort wird die Aufbewahrung von Primärdaten auf haltbaren und gesicherten Trägern für einen Zeitraum von zehn Jahren ausdrücklich empfohlen (vgl. auch American Psychological Association, 1994).

3 Sozialwissenschaftliche Datenarchive

Über die reine, von Einzelforschern zu leistende Archivierungsoption hinaus gelten in der Forschungsgemeinschaft Möglichkeiten des „data sharing“ als sinnvoll und unterstützenswert (Sieber, 1991). Nachprüfbarkeit und Replizierbarkeit von Untersuchungsergebnissen durch Dritte gehören zur allgemein anerkannten Definition von regelgerechter Forschung (Kaase, 1998; Wagner, 1999). Darüber hinaus ist der Mehrwert durch die Ermöglichung von Sekundäranalysen gerade in Anbetracht des häufig immensen Zeit- und Kostenaufwands zugrundeliegender Datenerhebungen beträchtlich (vgl. auch Procter, 1993).

Doch zugleich ist eine gewisse Zurückhaltung bezüglich der Bereitstellung eigener Forschungsdaten unübersehbar. Daten werden viel zu oft als „Eigentum“ betrachtet (vgl. Azar, 1999a, 1999b), das zu hüten und vor fremder Nutzung zu schützen ist. Die Vorbehalte gehen dabei häufig auf eine unklare Rechtslage bei der Überlassung von Forschungsdaten zurück. Die Verpflichtung potentieller Datennutzer zur Zitierung des Datengebers und ein verantwortungsvoller Umgang mit den überlassenen Daten muss daher als integraler Bestandteil des wissenschaftlichen Datenaustauschs verankert werden, der sich insbesondere durch eine klare zentrale Archivierungsstruktur sicherstellen lassen sollte.

Es gibt eine ganze Reihe von Einrichtungen und Initiativen zur Datenarchivierung und -bereitstellung in den Sozialwissenschaften. Als prominentestes Beispiel sei auf das United Kingdom Data Archive (UKDA) hingewiesen (vgl. Ageer, 2002). Mit dem Ziel, einen breiteren und besseren Umgang mit Forschungsdaten auch für Ausbildungszwecke zu erreichen, werden maschinenles-

bare Kopien von Datensätzen archiviert. Mehr als 4000 Sätze von Forschungsdaten stehen in diesem 1967 gegründeten Archiv mittlerweile zur Verfügung. In der Schweiz bemüht sich der Swiss Information and Data Archive Service for the Social Sciences (SIDOS) um die Archivierung sozialwissenschaftlicher Datensätze mit dem Ziel der Förderung von Sekundärforschung (vgl. Metral, 2002). Das Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln (ZA) ist das derzeit einzige öffentlich zugängliche Archiv für Daten der Sozialforschung in Deutschland (vgl. Jensen, 2003). Nach eigener Beschreibung ist seine Hauptaufgabe die Bereitstellung bereits vorliegender maschinenlesbarer Originaldaten, insbesondere Umfragedaten. Aus Übersee sei schließlich das Inter-university Consortium for Political and Social Research (ICPSR) genannt, das sich bereits seit 1962 um die Langzeitarchivierung sozialwissenschaftlicher Daten bemüht (vgl. Vardigan, 2003).

In die genannten Archive, die im übrigen über eine intensive Infrastruktur teils miteinander verbunden sind, haben auch psychologische Daten Eingang gefunden; allerdings werden sie nur selten und nicht systematisch erfasst und nachgewiesen. Zugleich stellt sich die Frage, inwieweit die vorwiegend auf Umfragedaten spezialisierten Archive die Aufnahme genuin psychologischer Daten unterschiedlichsten methodischen Ursprungs ermöglichen.

4 Das Psychologische Datenarchiv PsychData

Das Zentrum für Psychologische Information und Dokumentation (ZPID) – als überregionale, zentrale Dokumentations- und Informationseinrichtung für das Fach Psychologie in den deutschsprachigen Ländern – greift die Problematik der angemessenen Verwahrung von Primärdaten der empirisch-psychologischen Forschung in einem 2002 gestarteten Projekt auf, das von der Deutschen Forschungsgemeinschaft (DFG) gefördert wird. Ziel des Projekts ist der Aufbau des Psychologischen Datenarchivs PsychData und dessen Einbindung in das Dienstleistungsspektrum des ZPID (vgl. Zentrum für Psychologische Information und Dokumentation, 2002b).

Archivierungsgegenstand sind Primärdaten (in geringerem Umfang auch Ursprungsaufzeichnungen und abgeleitete Daten) aus allen Forschungsbereichen der deutschsprachigen Psychologie, insbesondere der Klinischen, der Entwicklungs-, der Pädagogischen, der Geronto- sowie der Arbeits- und Organisationspsychologie, die mit unterschiedlicher Methodologie (Umfrage, Test und Experiment) erhoben sein können. Primärdaten haben – im Vergleich zu Ursprungsaufzeichnungen und abgeleiteten Daten – den deutlichen Vorteil, dass sie alle relevanten Angaben der Versuchsteilnehmer relativ direkt widerspiegeln und zugleich eine handhabbare Form aufweisen.

Thematische Schwerpunkte der konzeptuellen Entwicklung des Psychologischen Datenarchivs PsychData liegen zunächst auf den Fragen der angemessenen, umfassenden Dokumentation von Forschungsdaten und der vor Datenalterung, -verlust und -manipulation geschützten Archivierung von Datenmaterial. Auf Basis der so gestalteten Datendokumentation und -archivierung steht die benutzerfreundliche und vertraglich klar geregelte Bereitstellung dieser Datenbestände für die Fachöffentlichkeit mit langfristigem Zeithorizont offen und damit die Ermöglichung vielfältigster Formen der Sekundäranalyse, wie etwa die Analyse neuer Fragestellungen, die Reanalyse mit anderen Methoden, die Optimierung von Parameterschätzungen und der Nachweis historischer Veränderungen (vgl. auch Hedrick, 1988). Die Weiterentwicklung von Forschungskultur und -ökonomie wird mithin nicht nur durch die Optimierung der Datenpflege im Primärforschungskontext angeregt, sondern zeitversetzt auch durch die Ermöglichung von Sekundärforschung vorangebracht.

5 Besonderheiten der Dokumentation von Primärdaten

Die Besonderheiten der Dokumentation von Primärdaten der empirisch-psychologischen Forschung leiten sich allein aus den Besonderheiten des archivierten Materials ab, wie sich an drei Diskussionspunkten verdeutlichen lässt:

Die Dokumentation von Primärdaten ist – stärker als jene anderer Archivierungsobjekte – notwendige Voraussetzung für deren Sinnhaftigkeit und Nutzbarkeit. Wird etwa ein wissenschaftlicher Zeitschriftenartikel oder Buchbeitrag nicht dokumentiert, bleibt dieser als Druck- oder elektronische Version aus sich selbst heraus lesbar und interpretierbar. Nicht dokumentierte Forschungsdaten – im Regelfall eine Ansammlung numerischen Materials in einer rechteckigen Matrix – bleiben hingegen informationsleere Ziffernfolgen.

Wenn die Dokumentation als unabdingbar erkannt wird, so stellt sich zum Zweiten die Frage, von wem sie geleistet werden kann. Schnell wird deutlich, dass die Dokumentation nur einem eingeschränkten Personenkreis – in der Regel dem Primärforscher oder einigen wenigen Personen aus dem unmittelbaren Forschungskontext – möglich ist. Unsaubere Dokumentationen der Daten im Forschungsprozess, das Ausscheiden verantwortlicher Forscher aus dem Projekt, Tücken der EDV oder der menschlichen Erinnerungsleistung führen immer wieder dazu, dass die Bedeutung einzelner Datenpunkte nicht mehr rekonstruiert werden kann. Diese Datenpunkte sind für die Nachnutzung auf Dauer verloren.

Eine dritte Besonderheit der Dokumentation von Primärdaten der empirisch-psychologischen Forschung liegt darin begründet, dass für eine Nutzung des Datensatzes allein das Verständnis der Datenmatrix ungenügend bleibt. Die bei der Datenerhebung eingesetzten Instrumente, die zugrundeliegende Studie

einschließlich ihres konzeptuellen Backgrounds, die untersuchte Stichprobe als Ausschnitt der avisierten Population und die Besonderheiten der Erhebungssituation stellen wesentliche Informationseinheiten dar, ohne die Primärdaten nur begrenzt zu nutzen sind. Dass die Datensätze außerdem bibliographisch erschlossen werden sollten, versteht sich von selbst und fügt der Dokumentation eine weitere Ebene hinzu.

6 Projektspezifische Metadaten

Die Dokumentation eines jeden Archivierungsobjekts verlangt die Festlegung eines Satzes von Beschreibungsfeldern: Die projektspezifischen Metadaten. Diese müssen natürlich die allgemeinen Funktionalitäten von Metadaten erfüllen: Unterstützung des Nutzers bei der Recherche, Bewertungshilfe hinsichtlich der Eignung für eigene Nutzungszwecke und Beschreibung der Zugriffsmöglichkeiten. Vor dem Hintergrund der diskutierten Besonderheiten der Archivierungsobjekte müssen die Metadaten vor allem aber die Primärdaten interpretierbar und damit nutzbar machen. Damit im Einklang steht die Definition des ISO/IEC Standard 11179, wonach Metadaten zu verstehen sind als „information and documentation which makes data sets understandable and sharable for users“ (Gillman, 1997).

Die Entwicklung eines Sets von Metadaten im Rahmen des Psychologischen Datenarchivs PsychData bewegt sich in einem Spannungsfeld aus Forderungen verschiedenster Couleur. Vor dem Hintergrund internationaler Bestrebungen, über verschiedene Archivierungskontexte hinweg standardisierte Beschreibungsmodalitäten festzuschreiben und so die Interoperabilität verschiedenster Archivierungsobjekte zu ermöglichen, sind zunächst möglichst allgemeine Metadatenstandards zu integrieren. Der als fachübergreifender Metadatenstandard entwickelte Dublin Core (Dublin Core Metadata Initiative, 2001) konnte in das projektspezifische Metadatenset implementiert werden. Die Funktionalitäten des Psychologischen Datenarchivs soll in diesem Zusammenhang durch den integrierten Zugriff auf forschungsrelevante Informationseinheiten wie Fachdatenbanken oder Volltextserver unterstützt werden, deren Einbindung bereits bei der Entwicklung der projektspezifischen Metadaten – etwa durch die Aufnahme zentraler Literaturverweise und Testpublikationen oder die an Standards orientierte Verschlagwortung (vgl. Zentrum für Psychologische Information und Dokumentation, 2002a) – zu bedenken ist. Neben allgemeinen sind bereichsspezifische – sprich sozialwissenschaftliche – Dokumentationsstandards nicht zu vernachlässigen. Die Entwicklung des Metadatenystems vollzog sich daher unter besonderer Beachtung der Document Type Definition der Data Documentation Initiative (vgl. Vardigan, 2002). Gerade in Auseinandersetzung mit diesem sozialwissenschaftlichen Standard traten die besonderen Erfordernisse der

Dokumentation empirisch-psychologischen Forschung deutlich zu Tage. Eine Ergänzung der DDI-Kategorien war dabei in einigen Bereichen ebenso erforderlich wie in anderen Sektoren eine Kürzung der Beschreibungskategorien möglich. Darüber hinaus ist die Belastung der Datengeber – die als einzige über die relevante Information verfügen – möglichst gering zu halten und die Koordination und Dokumentation des internen Arbeitsablaufs sicherzustellen.

Auf Basis verschiedener Entwicklungsschritte resultierte ein Set von knapp 70, teils optionalen Metadatenfeldern. Die im Vergleich zu vielen anderen Archivierungskontexten recht umfangreichen Metadaten lassen sich in interne Verarbeitungsdaten, formale Studienbeschreibung, inhaltliche Studienbeschreibung, Beschreibung der Dateien und Materialien und das Kodebuch unterteilen.

Im Rahmen der internen Verarbeitungsdaten werden zunächst Angaben zum ZPID-Bearbeiter, Bearbeitungsstatus und zur Ablage selbst vermerkt. Aufgrund der engen Zusammenarbeit mit dem Datengeber ist zusätzlich die Koordination dieser Interaktionen erforderlich, wie etwa der Vermerk auf relevante Kontakte und wesentliche Termine. Einige dieser Angaben (z. B. Ablageinformationen, externer Bearbeiter) nehmen als Kategorien des Dublin Core eine hervorgehobene Position ein.

Die formale Studienbeschreibung enthält typische Metainformationen über den Datensatz als Ganzes. Neben Angaben wie Autor und Titel des Datensatzes sind Angaben zum Funding, Verantwortlichkeit für die Datenerhebung und Copyright zu vermerken. Mit näherem Bezug zum Datensatz werden diese durch Datenstatus, Anonymisierung und Kodierung der Forschungsform ergänzt. Eine Verschlagwortung des Datensatzes nach standardisierter Terminologie bereitet wertvolle Rechercheoptionen vor. Relevante Veröffentlichungen, die sich direkt auf den Datensatz beziehen oder einschlägig über das Forschungsgebiet informieren, runden das Angebot an formalen und bibliographischen Angaben ab. In diesem Segment befinden sich zentrale bibliographische Angaben des Dublin Core.

Kernstück des Metadatenatzes und zugleich das von den gängigen Dokumentationsstandards wie Dublin Core und DDI am stärksten abweichende Segment formiert die inhaltliche Studienbeschreibung. Angaben zur Studie, zu den forschungsleitenden Hypothesen, vor allem aber zu den Erhebungsinstrumenten, zur Stichprobe, Population und zum Erhebungsdesign führen tief in das Studienumfeld des Datensatzes ein. Ein näherer Einstieg in die Datenmatrix wird durch die konzeptuelle Erschließung und Gruppierung der Variablen geleistet.

Wenngleich bislang von einer Datenmatrix als Archivierungsobjekt gesprochen wurde, so werden neben dieser in der Praxis oftmals weitere Dateien oder Materialien beigelegt, wie etwa Exemplare des Erhebungsinstruments, Instruktionen für Versuchspersonen oder Versuchsleiter, Textdokumente mit Detailinformationen über die Studie oder auch Quellcodes bei computergestützten Er-

hebungen. Dem grundlegenden Archivierungsgedanken folgend werden alle eingereichten Bestandteile in unveränderter Form in dem Psychologischen Datenarchiv PsychData verwahrt. Sollen die Primärdaten für eine Sekundärnutzung bereitgestellt werden, ist eine reine Archivierung in der eingehenden Form jedoch unzureichend. Vielmehr sind umfangreiche Prüfungen und oftmals auch Überarbeitungen der eingereichten Materialien erforderlich. Die Ablage der Primärdatenmatrix – einleitend als wesentlicher Schritt des Forschungsprozesses konzipiert – wird in der Praxis häufig nach Durchlaufen des Forschungsprozesses an das Psychologische Datenarchiv übertragen. Zugleich sind die zu meist im auswertungsrelevanten Format eingereichten Daten in nicht-proprietäre Formate zu wandeln und von programmspezifischen Besonderheiten zu bereinigen. Insbesondere die Vollständigkeit des Kodierplans und seine Übereinstimmung mit dem Datendokument sind kritisch zu prüfen. Selbstverständlich sind gerade solche Schritte sorgfältig zu dokumentieren. Auch ist eine begründete Entscheidung dahingehend zu treffen, welche weiteren Informationen ein externer Nutzer zur Sekundäranalyse des Datensatzes benötigt.

Das Kodebuch schließlich beschreibt das Archivierungsobjekt auf einer datennahen Ebene. Zu jeder Primärvariablen werden Variablenlabel, Variablenbeschreibung, Itemtext (sofern vorhanden) und insbesondere die Kodierung valider und fehlender Werte mitgeteilt. Anders als bei der DDI, bei der das Kodebuch einen integralen Bestandteil des Metadatensets formiert, wird dem Kodebuch im Psychologischen Datenarchiv eine Sonderrolle zugeschrieben. Die hier auf Variablenebene abgelegte Information bleibt – anders als etwa im Bereich der Sozialwissenschaften – ohne Rückgriff auf die in der inhaltlichen Studienbeschreibung angeführten Konzepte für ein Retrieval von eher geringem Wert.

Alle Metadaten werden in möglichst standardisierter Form abgelegt, wenn gleich erste Erfahrungen zeigen, dass die Vielfalt psychologischer Erhebungsansätze und -methoden eine vollständige Standardisierung im Sinne festgeschriebener Auswahloptionen bei vielen Metadatenfeldern verunmöglicht. Fernziel ist dennoch die Entwicklung kontrollierter Vokabulare für die meisten Beschreibungsfelder.

Das entwickelte Metadatenchema befindet sich derzeit in einer ersten Prüfung. Studien aus unterschiedlichen Inhaltsbereichen und mit unterschiedlichem methodischem Zugang werden unter Rückgriff auf das Metadatenchema dokumentiert. Die daraus resultierenden Erfahrungen dienen zur Prüfung und gegebenenfalls zur Modifikation der ausgewählten Beschreibungsfelder sowie zur fortschreitenden Standardisierung der Beschreibungssprache. Im Laufe des Jahres 2004 wird die Datenbank, die auf dem projektspezifischen Metadatenchema aufsetzt, über die Projekthomepage <http://psychdata.zpid.de> für Nutzer bereitgestellt werden.

7 Ausblick

Die Unterstützung von Wissenschaftlern aus dem Bereich der empirisch-psychologischen Forschung steht im Zentrum des Psychologischen Datenarchivs PsychData: Einerseits die Unterstützung von Primärforschern, die bei der Aufbereitung und Ablage ihrer Primärdaten auf ein professionelles Angebot zurückgreifen können und andererseits die Unterstützung von Sekundärforschern, die Zeit- und Kostenaufwand für erneute Datenerhebungen durch den Zugriff auf bereits vorhandenes Datenmaterial einsparen können.

Ein Angebot wie das dargestellte ist jedoch – und darauf ist am Ende dieser Ausführungen in aller Deutlichkeit hinzuweisen – im Bereich der empirisch-psychologischen Forschungsgemeinschaft noch wenig etabliert. Im Primär- und Sekundärforschungskontext ist daher nicht nur die konkrete, datensatzbezogene Arbeit zu leisten; dieser muss vielmehr eine ausführliche Information der Forschenden vorangehen.

Das Ziel, Daten zum Erleben und Verhalten von Menschen an zentraler Stelle zu dokumentieren, zu archivieren und für Sekundäranalysen bereitzuhalten, lässt sich – fernab aller Diskussion um Forschungskultur und -ökonomie – nicht zuletzt auch mit den Belastungen der Versuchsteilnehmer begründen, die mit ihrem Engagement für die Wissenschaft eine optimale Ausschöpfung der bereitgestellten Daten verdienen.

Literatur

- Ageer, L. B. (2002). About the UK Data Archive. Verfügbar unter URL: <http://www.data-archive.ac.uk/home/theArchive.asp#Top> (2003-04-08).
- American Psychological Association (1994). Publication manual of the American Psychological Association (4th ed.). Washington, DC: American Psychological Association.
- Azar, B. (1999a). Psychology needs to develop mechanisms for data sharing. There's little common ground on how, when and whether data should be shared. APA Monitor Online, Vol. 30, Nr. 8. Verfügbar unter URL: <http://www.apa.org/monitor/sep99/sc2.html> (2003-04-08).
- Azar, B. (1999b). Will you be forced to share your data? Funding agencies ponder new policies that call for grantees to archive and share their research data. APA Monitor Online, Vol. 30, Nr. 8. Verfügbar unter URL: <http://www.apa.org/monitor/sep99/data.html> (2003-04-08).
- Deutsche Forschungsgemeinschaft (1998). Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Verfügbar unter URL:

- http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf (2003-04-08).
- Dublin Core Metadata Initiative (2001). Dublin Core Metadata Initiative (DCMI) Home Page. Verfügbar unter URL: <http://dublincore.org/> (2003-04-08).
- Erdfelder, E. (1994). Erzeugung und Verwendung empirischer Daten. In T. Herrmann & W. H. Tack (Hrsg.), *Enzyklopädie der Psychologie. Themenbereich B Methodologie und Methoden, Serie I Forschungsmethoden der Psychologie*, Bd. 1 Methodologische Grundlagen der Psychologie (S. 47-97). Göttingen: Hogrefe.
- Gillman, D. W. (1997). ISO/IEC Standard 11179. Specification and standardization of data elements. Part 1 Framework for the specification and standardization of data elements. Verfügbar unter URL: <http://www.ssd.rl.ac.uk/ccsdsp2/Fra97/11179-1.txt> (2003-04-08).
- Hedrick, T. E. (1988). Justifications for the sharing of social science data. *Law and Human Behavior*, 12, 163-171.
- Jensen, U. (2003). Zentralarchiv für Empirische Sozialforschung, Universität zu Köln. Verfügbar unter URL: <http://www.gesis.org/ZA/index.htm> (2003-04-08).
- Kaase, M. (1998). Fälschungen von Daten und Resultaten? Maßnahmen zur Vorsorge. *Soziologie, ohne Bandangabe*, Heft 2, S. 95-96.
- Metral, G. (2002). Was macht SIDOS? Verfügbar unter URL: <http://www.sidos.ch/about/default.asp?lang=d&menu=1> (2003-04-08).
- Selg, H., Klapprott, J. & Kamenz, R. (1992). *Grundriß der Psychologie*. Bd. 3 *Forschungsmethoden der Psychologie*. Stuttgart: Kohlhammer.
- Sieber, J. E. (Ed.). (1991). *Sharing social science data. Advantages and challenges*. Thousand Oaks, CA: Sage.
- Procter, M. (1993). Analyzing other researchers' data. In N. Gilbert (Ed.), *Researching social life* (pp. 255-269). London: Sage.
- Vardigan, M. (2002). About the organisation. Verfügbar unter URL: <http://www.icpsr.umich.edu/DDI/ORG/index.html> (2003-04-08).
- Vardigan, M. (2003). About ICPSR. Verfügbar unter URL: <http://www.icpsr.umich.edu/org/index.html> (2003-04-08).
- Wagner, G. (1999). Ziele und Unabhängigkeit der Wissenschaft sind Instrumente eines effektiven Datenschutzes. In R. Hamm & K. P. Möller (Hrsg.), *Datenschutz und Forschung* (S. 14-20). Baden-Baden: Nomos.
- Zentrum für Psychologische Information und Dokumentation (Hrsg.). (2002a). *PSYINDEX Terms: Deskriptoren/Subject Terms zu den Datenbanken PSYINDEX und PSYTKOM* (6., aktualisierte Aufl.). Trier: Zentrum für Psychologische Information und Dokumentation, Universität Trier.

Zentrum für Psychologische Information und Dokumentation (2002b). ZPID Leistungsspektrum. Verfügbar unter
URL: <ftp://ftp.zpid.de/pub/profil/spektrum.pdf> (2003-04-08).

Kontakt

Jutta von Maurice, Dipl.-Psych.
Zentrum für Psychologische Information und Dokumentation (ZPID)
Universität Trier
Universitätsring 15
54296 Trier
Tel.: +(49)-(0)651-201-2867
Fax: +(49)-(0)651-201-2071
E-Mail: maurice@zpid.de
Internet: www.zpid.de

Maßnahmen zur Förderung der Informationskompetenz durch den Fachinformationsanbieter ZPID

Astrid Nechvátal

ZPID

With increased access to the Internet barriers to researching scientific information have fallen sharply. Improvements to user interfaces and inexpensive prices should contribute to further widespread use. However, usage is not as prolific as it should be, even among researchers. Scientists as well as students often use simple, unspecific Internet search engines instead of quality databases to research information. Results of the SFS study "Use of Electronic Information in Higher Education" show that the use of electronic sources of scientific information is not sufficiently anchored in university education, information exchange among students is inadequate, and information resources are experienced as too complex. Information literacy and knowledge management skills are not recognized as key qualifications in higher education, and cooperation among various information providers such as national information centers, academic and specialized libraries, and university instructors is viewed as deficient. Results of the SFS study as well as similar studies stress the importance of cooperation among various information agents in the conceptualization, development, and testing of training modules. As a national institute for psychology information, ZPID has responded to this demand by providing online training materials for self-study and use by university instructors.

Die Nutzung elektronischer wissenschaftlicher Information erweist sich in der Hochschulausbildung bislang als defizitär. Im folgenden Beitrag werden bestehende Defizite aufgezeigt und die daraus resultierenden Ansatzpunkte und konkreten Maßnahmen zur Förderung von Informationskompetenz durch das Zentrum für Psychologische Information und Dokumentation (ZPID) als Fachinformationseinrichtung dargestellt. Neben allgemeinen konzeptuellen Überlegungen zu entsprechenden Lehr- und Lernmodulen wird insbesondere die spezielle Position einer Fachinformationseinrichtung im Kontext der Förderung von Informationskompetenz beleuchtet.

Ausgangslage

Seit der Etablierung des Internets sind die Barrieren für den Zugang zu Fachinformationen gefallen. Im Wissenschaftsbereich geht man von einer nahezu deckenden Vollversorgung aus. Etwa 95 % der Studierenden können auf einen universitären Computerarbeitsplatz mit Internetanschluss zurückgreifen und gut 70 % der Studierenden verfügen zudem über einen eigenen Computer mit Internetanschluss (vgl. Klatt, Gavrilidis, Kleinsimlinghaus & Feldmann, 2001b). Fortschritte in der Bedienerfahrung und günstige Kosten sollten ein übriges tun, um die Nutzung elektronischer wissenschaftlicher Information voranzutreiben. Trotz der günstigen Ausgangslage bleibt dies jedoch selbst im Rahmen von Forschungsvorhaben weit hinter der sachlich gebotenen Notwendigkeit zurück. Wissenschaftler wie Studierende verwenden oft simple, unspezifische Internet-Suchmaschinen, statt qualitativ hochwertige Datenbanken zu befragen.

Defizite im Zugriff auf Fachinformationen wurden im Jahr 2001 von der Sozialforschungsstelle Dortmund (SFS) in Zusammenarbeit mit dem Lehrstuhl für Soziologie der Universität Dortmund und der Gesellschaft für Angewandte Unternehmensforschung und Sozialstatistik mbH (GAUS) im Auftrag des Bundesministeriums für Bildung und Forschung (BMB+F) im Rahmen einer repräsentativen Studie zur „Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung - Barrieren und Potenziale der innovativen Medienutzung im Lernalltag der Hochschulen“ eindrucksvoll belegt.

Zielsetzung der Untersuchung war es, die Informationskompetenz von Studierenden und Hochschullehrenden zu ermitteln, die damit verbundenen Potenziale auszuloten und darauf aufbauend Maßnahmen vorzuschlagen, die zu einer Verbesserung der Nutzung elektronischer wissenschaftlicher Information beitragen können. Die Befragung richtete sich an drei Zielgruppen: Dekanate (n = 353), Studierende aus den Fachbereichen Chemie, Informatik, Mathematik, Physik, Elektrotechnik, Bauingenieurwesen, Maschinenbau, Psychologie, Pädagogik/Erziehungswissenschaften, Soziologie/Sozialwissenschaften (n = 2956) und Hochschullehrende (n = 777).

Als zentrales Ergebnis der insbesondere auf institutionelle Aspekte und institutionelle Kontextfaktoren bezogenen Befragung von Dekanaten lässt sich festhalten, „dass eine formale Verankerung der Nutzung elektronischer wissenschaftlicher Information im Studium noch nicht nachhaltig stattgefunden hat. Korrespondierend dazu ist die personelle Ausstattung zur Integration elektronischer Informationsmedien weitgehend noch nicht ausreichend“ (S. 6, Klatt et al., 2001b).

Die Befragung der Studierenden zeigte, „dass die Informationskompetenz der meisten Studierenden zur Nutzung elektronischer wissenschaftlicher Information unzureichend ist. Die Studierenden empfinden das Angebot dieser Informationen als unübersichtlich. Sie sehen sich außerstande, die Qualität und

Bedeutung der mit elektronischer Recherche verbundenen Ergebnisse zu bewerten Statt das gesamte Spektrum fachspezifischer elektronischer wissenschaftlicher Medien zu nutzen, beschränken sich die Studierenden häufig auf das 'Browsen' im Internet mit Hilfe freier Suchmaschinen wie Lycos oder Web.de. Ob sie dabei auf wertvolle und hilfreiche Informationen stoßen, bleibt ihnen verschlossen, weil sie ihre Kenntnisse im Umgang mit den neuen Medien nicht systematisch im Rahmen ihres Studiums, sondern im Selbstlernverfahren erworben haben“ (S. 4, Klatt, Gavrilidis, Kleinsimlinghaus & Feldmann, 2001a).

„Dies führt(e) zu Halbwissen und dazu, dass die Bedürfnisse der dilettierend fortgeschrittenen und ebenso fortschreitenden Studierenden von Einführungsveranstaltungen zu elektronischen wissenschaftlichen Informationen nicht mehr hinreichend getroffen werden. Dem entspricht, dass die Studierenden in der Regel mit zunehmender Studiendauer zwar immer mehr, aber offensichtlich unsystematisch und unvollständig Kenntnisse zur Nutzung elektronischer wissenschaftlicher Informationen erlangen und es korrespondiert mit der Präferenz der Studierenden für übersichtliche und vergleichsweise einfache Formen und Angebote der elektronischen wissenschaftlichen Informationen“ (S. 13, Klatt et al., 2001b).

Die Ergebnisse der Hochschullehrendenbefragung „lassen insgesamt auch eine mangelnde Qualifizierung bei manchem Lehrenden erkennen, sich im Dickicht elektronischer wissenschaftlicher Information wirklich effizient zurechtzufinden. Auch etliche Dozent/innen verschenken durch vorwiegend autodidaktisch erworbene Kompetenzen in bezug auf die Nutzung elektronischer wissenschaftlicher Information die Chance des gezielten Zugangs zu systematischer, bewerteter wissenschaftlicher Information zugunsten leicht zugänglicher und zufallsanfälliger Informationswege“ (S. 4, Klatt et al., 2001a).

Forderungen

Aufbauend auf Defiziten und Potenzialen wurden Ziele und Maßnahmen definiert, die zu einer Verbesserung der Nutzung elektronischer wissenschaftlicher Information beitragen können.

Die grundsätzlichsste Forderung betrifft dabei die inhaltliche Definition und Verankerung der Schlüsselqualifikation Informationskompetenz im universitären Kontext. So sollen die Anforderungen an die Informationskompetenz in der Hochschulausbildung zwischen Unternehmen und Hochschulen abgestimmt werden, um die Informationskompetenz der Studierenden gemeinsam anwendungsorientiert aufzubauen. Auch die Förderung und Institutionalisierung des Wissensaustausches unter den Studierenden bei der Nutzung und Verwendung elektronisch bereitgestellter Fachinformation z. B. über Tutorien wird als Maß-

nahme definiert, ebenso wie die Minderung der beklagten Unübersichtlichkeit durch Bereitstellung geprüfter Angebote elektronischer wissenschaftlicher Information gemäß den fachlichen Anforderungen durch die Fachbereiche bzw. Fakultäten.

Eine weitere relevante Forderung bezieht sich auf die Verankerung der Nutzung elektronischer wissenschaftlicher Information als integralen Bestandteil der Lehre. Besondere Bedeutung wird dabei der Etablierung der Informationskompetenzentwicklung im normalen Studienalltag beigemessen. Hochschullehrende als Vorbilder, Multiplikatoren und diejenigen mit den genauesten Kenntnissen der Studieninhalte, haben die Möglichkeit, die Nutzung elektronischer wissenschaftlicher Information anhand konkreter Studieninhalte gewinnbringend aufzuzeigen und nachhaltig zu implementieren.

Um die Grundlagen hierfür zu schaffen, müssen Lehrveranstaltungen oder Lehr-Lernmodule konzipiert, erstellt und praktisch erprobt werden. Damit dies kosten- und zeitsparend, studienbereichsspezifisch und umfassend geleistet werden kann, bedarf es hochschulinterner und hochschulübergreifender Kooperationen und Netzwerke.

„Die Konzipierung und Durchführung solcher Lehrveranstaltungen oder Lehr-Lernmodule sind zeit- und personalintensiv. Sie bedürfen desto eher einer Kooperation verschiedener Akteur/-innen aus Lehre, Universitäts- und auch Fachbereichsbibliotheken, je spezifischer die zu vermittelnden Kompetenzen zur intelligenten Nutzung elektronischer wissenschaftlicher Information sein sollen“ (S. 236, Klatt et al., 2001b). „Jeder von uns, der mit Internetangeboten zu tun hat, weiß, dass die Entwicklung eine Sache ist, die Pflege und Weiterentwicklung aber ab einem bestimmten Punkt nicht mehr zu bewältigen ist und nur kooperatives Vorgehen eine Garantie für Qualität und Nachhaltigkeit ist“ (S. 5, Schefczik, 2001).

Die Zusammenarbeit von Fachinformationszentren, Universitäts-/ Fakultätsbibliotheken und Hochschullehrenden zur Förderung der Informationskompetenz wird daher explizit und nachdrücklich gefordert.

Inhaltlich werden bei der Vermittlung von Informationskompetenz die Schwerpunkte vor allem auf allgemeinen Strategien, universalen Strukturen sowie grundsätzlichen Erkenntnis- und Bewertungsschemen liegen, um der Flexibilität einer sich stetig wandelnden Informationslandschaft gerecht werden zu können.

Ein wichtiger Punkt ist nicht zuletzt die Vermittlung von spezifischem Wissen um - und Strategien zur Nutzung und Bewertung von - Fachinformationseinrichtungen, Fachdatenbanken und Fachportalen. An dieser Stelle ist das ZPID gefragt als überregionale, zentrale Fachinformationseinrichtung für das Fach Psychologie mit ihren Datenbanken, wie etwa PSYINDEX oder PSTKOM, psychologischem Fachportal und spezifischer Suchmaschine.

Konzeption

Die Bereitstellung und Vermittlung fachspezifischer Informationen zu den Nutzungsmöglichkeiten der psychologischen Fachinformationseinrichtung ZPID stehen im Mittelpunkt der Konzeption eines Online-Lehr-Lernmoduls. Das Modul umfasst derzeit einen Pool von 87 Folien, die in einer Power-Point-Präsentation zusammengefasst sind und auf der ZPID-Homepage zum freien Download bereitgestellt werden (Zentrum für Psychologische Information und Dokumentation, 2003).

Die Präsentation berücksichtigt mehrere, in unterschiedlichen Lern- und Lehrsituationen relevante Themenbereiche. Neben allgemeinen Informationen über das ZPID als psychologische Fachinformationseinrichtung und Nutzungshinweisen zur Präsentation werden Bausteine zu den verschiedenen inhaltlichen Arbeitsbereichen des ZPID angeboten. Die thematisierten Segmente umfassen die im psychologischen universitären Kontext allgemein verwendeten Datenbanken PSYINDEX und PSYTKOM sowie weniger bekannte kostenfreie Datenbanken, wie etwa die Diplomarbeitsdatenbank oder das Verzeichnis der Testanbieter. Weitere Blöcke reichen von der Einführung in die psychologische Suchmaschine PsychSpider über Informationen zu Publikationen, wie dem Nachschlagewerk PSYINDEX Terms oder dem elektronischen Verzeichnis der Testverfahren, bis hin zu den Thematiken Auftragsrecherchen, Informations- und Dokumentationsforschung, Drittmittelprojekte und Benutzerschulungen.

Da das allgemein zugängliche Lehr-Lernmodul auch unerfahrenen Nutzern die Möglichkeiten des ZPID-gestützten Zugriffs auf relevante Fachinformationen näher erschließt, sollte es zur Förderung der Informationskompetenz beitragen. Dabei wird Informationskompetenz verstanden als die Fähigkeit zu erkennen, wann Information benötigt wird, sie zu finden, zu evaluieren und effektiv zu nutzen. Umfassende Kenntnis über Zugänglichkeit und Nutzen der für den Studienbereich relevanten Informationsquellen ist dabei ein wichtiger Kernpunkt.

Die entwickelte ZPID-Präsentation kann interessierten Endnutzern unmittelbar als Lernmodul dienen. In dieser unmittelbaren Verwendung sollte das bereitgestellte Modul etwa zur Optimierung der Informationskompetenz von Studierenden beitragen. Auch Hochschullehrende könnten sich mit Hilfe der Unterlagen auf einfache Weise über aktuelle Entwicklungen der Fachinformationseinrichtung ZPID informieren.

Bedeutsamer als diese Verwendung als Lernmodul ist jedoch die Nutzung als Lehrmodul durch Multiplikatoren im Netzwerk. Hochschullehrende und Bibliotheken sind gehalten, sich in unterschiedlichen Studienphasen um die Optimierung der Informationskompetenz der Studierenden zu bemühen. Mit Hilfe des entwickelten Moduls lässt sich mit geringem Aufwand Schulungsmaterial generieren, dass die spezifische Situation des Adressatenkreises berücksichtigt.

Um das Lehr-Lernmodul an die konkreten Bedürfnisse und Erfordernisse von Multiplikatoren oder Endnutzern anpassen zu können, wurden bei der Konzeption drei Hauptachsen verfolgt: Flexibilität, Informationsvermittlung „vertikal“ und Bedienungskomfort.

Flexibilität

Ein flexibles Lehr-Lernmodul ermöglicht dem Nutzer eine Anpassung vorgegebener Produkte an individuelle Erfordernisse. Die Forderung nach Flexibilität spiegelt sich in dem ZPID-Angebot in den Punkten Systemvoraussetzungen, Bildschirmpräsentation und Themenauswahl wider. Das ungewöhnlich hohe Maß an Entscheidungs- und Adaptationsmöglichkeiten durch den Nutzer umfasst Aspekte wie technische Voraussetzungen, zeitliche Rahmenbedingungen, spezifische Zielgruppen und Themen.

Systemvoraussetzungen

Da auch diejenigen als Multiplikatoren angesprochen werden sollen, die selbst noch die ein oder andere Hemmschwelle zur Nutzung bestimmter technischer Systeme haben, ist die Wahl der Datenbereitstellung ein wichtiger Aspekt. Zudem muss je nach Einsatzort von sehr unterschiedlicher technischer Ausstattung ausgegangen werden. Nicht jeder Seminarraum verfügt über die Möglichkeit, Präsentationen online vorzuführen, bzw. über einen entsprechend einsatzfähigen PC.

Die Entscheidung, die Lehr-Lernmodule als animierte PowerPoint-Präsentationsfolien zu erstellen und zusätzlich als Pdf-Dateien verfügbar zu machen, basiert auf Praxiserfahrungen und entsprechenden Rückmeldungen. Zurzeit werden die Daten auf der Internetplattform frei zugänglich als PowerPoint 2000, PowerPoint 97, Pdf-Datei (Folienansicht) sowie Pdf-Datei (inklusive Notizseiten) angeboten (Zentrum für Psychologische Information und Dokumentation, 2003). Unter Umständen wird sich hier in den nächsten Jahren die Bereitstellung in anderer Form als sinnvoll erweisen.

Bildschirmpräsentation

Um dem Nutzer einen flexiblen Einsatz des ZPID-Moduls zu ermöglichen, können aufgrund der angebotenen Formate unterschiedliche Präsentationsarten gewählt werden. Die Vorlagen sind nutzbar als Overhead-Präsentation, als Online- oder Offline-PowerPoint-Präsentation, mit oder ohne Animation, mit voreingestellter Anzeigedauer oder manuell via Mausclick oder Tastatur. Die präsenta-

tionstechnischen Variablen sind frei wählbar, Animation, voreingestellte Anzeigendauer und Folienübergänge individuell veränderbar.

Erfahrungswerte zeigen, dass der Umgang mit dieser Anwendung oftmals auf einer elementaren Ebene stattfindet und das Nutzerwissen, gerade was Präsentationsformen und -möglichkeiten anbelangt, überaus heterogen ist. Da die Materialien auch diejenigen als Multiplikatoren ansprechen sollen, denen der Umgang mit den neuen Medien und Präsentationsformen nicht in allen Facetten bekannt ist, wurden Bearbeitungshinweise eingebaut, die auch den eher Ungeübten die Nutzung und Veränderung spezifischer Parameter der PowerPoint-Präsentation ermöglichen.

Themenauswahl

Die Möglichkeit der Differenzierung des Schulungsangebots nach den aktuellen Informationsbedürfnissen stellte eine grundlegende Anforderung dar. Je nach Themenschwerpunkt, Zielgruppe und Zeitumfang können daher die Folien individuell aus einem Folienpool zusammengestellt werden. Eine themenspezifische Farbmarkierung auf den Folien ermöglicht die gezielte und komfortable Auswahl der gewünschten Elemente und deren Kombination zu einzelnen Lehr-Lernmodulen. Eine Hinweisfolie im ersten Abschnitt der Präsentation vermittelt dem Nutzer das entsprechende Wissen (vgl. Abb. 1).

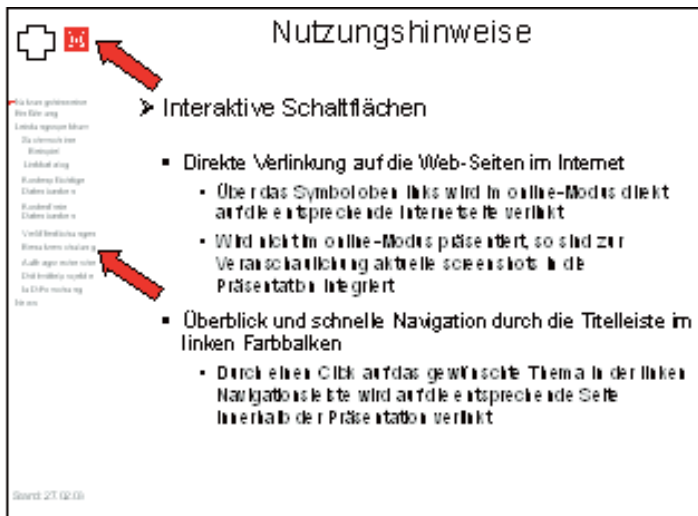


Abb. 1: Nutzungshinweise zur themenspezifischen Farbmarkierung

Inwieweit diese Möglichkeit tatsächlich genutzt wird bzw. als zusätzlicher Aufwand auf Vorbehalte stößt, werden die Rückmeldungen zeigen. Sicherlich ist hier eine grundsätzliche Problematik angesiedelt. Zum einen werden spezifische, hochangepasste und individuelle Lösungen gefordert – auf der anderen Seite zusätzlicher Aufwand gescheut und Fertiglösungen präferiert. Zu Bedenken ist jedoch, dass nur durch eine enge Anbindung an die subjektiven und fachbezogenen Handlungs- und Informationsbedürfnisse das Motivationspotential, das in diesem Augenblick vorhanden ist, effektiv und nachhaltig genutzt werden kann.

Die Sinnhaftigkeit der Beschäftigung mit Information und die Notwendigkeit Informationskompetenz zu entwickeln, sowie ein Verständnis für die Bedeutung von Informationskompetenz ist nur in der konkreten Verknüpfung mit einem aktuellen Informationsbedürfnis, z. B. im Zusammenhang mit Facharbeiten oder Diplomarbeiten nachhaltig möglich. „Am richtigen Ort zum richtigen Zeitpunkt, dann, wenn die Kunden der Bibliothek dies benötigen, können so wesentliche Elemente von Informationskompetenz vermittelt werden“ (S. 3, Hapke, 2002).

Benno Homann spricht im August 2001 im Rahmen eines Vortrages über Schwierigkeiten und neue Ansätze der Benutzerschulung in Deutschland von Einzelveranstaltungen als Lernmodule, die über ein kleines Curriculum miteinander verknüpft sind, sowie von Einzelkompetenzen und Lernzielen, die auf die einzelnen Veranstaltungen verteilt werden und sich in ihren inhaltlichen Schwerpunkten unterscheiden. „Sie erleichtern die didaktische Reduktion von Inhalten und ermöglichen ein teilweise hierarchisch gestuftes Schulungsangebot, wie es in Heidelberg mit der Unterscheidung in Grundlagenveranstaltungen und Aufbauschulungen realisiert wurde“ (S. 5, Homann, 2001).

Zudem verhindert die spezifische Aufteilung in themenspezifische Module eine inhaltliche Überlastung sowie inhaltlich überflüssige Redundanzen. Gleichzeitige können gezielt hilfreiche Redundanzen zur Festigung der Kompetenz etabliert werden.


Informationsvermittlung „vertikal“

Ein weiterer wesentlicher Aspekt bei der Konzeption des ZPID-Moduls war neben einer einfachen übersichtlichen Strukturierung die Vermittlung von Basis- und Hintergrundwissen.

„Vertikale Informationsvermittlung“ beschreibt den Ansatz, Informationen zusätzlich in der Tiefe zu verdichten und nicht nur möglichst breit zu fächern.

Insbesondere unter dem Aspekt der Evaluation, der Bewertung und Beurteilung von Informationen wurden zusätzlich zu den begrenzten Informationskapazitäten der Folien Hintergrundinformationen auf den Notizseiten platziert.

Kriterien wie Verlässlichkeit, Aktualität und Vollständigkeit sind bedeutsame Faktoren zur Beurteilung. Daher befinden sich in den Unterlagen auf den Notizseiten unter anderem Hinweise zu Aktualisierungszeiträumen von Angeboten, zur Menge des Anwachsens pro Zeitraum, zu Suchstrategien, zur Aufbereitung der Inhalte und der fachlichen Qualität. Abbildung 2 zeigt das Zusammenspiel von Folie und Notizseite am Beispiel einer Informationseinheit zur Datenbank PSYINDEX.



Fachkommission
Medizinische
Fachinformatiker
für Gesundheits-
Informationswissenschaften
F + I
Leitung:
Prof. Dr. Grottel
Telefon:
+49 (0) 30 3354-2200
Fax:
+49 (0) 30 3354-2201
E-Mail:
f+i@fki.fhn.de

PSYINDEX

➤ Literatur + Audiovisuelle Medien

- Mehr als 170.000 Nachweise (Bibliographische Angaben, Kurzreferate, Schlagworte)
- Jährlicher Zuwachs von über 7.000 Nachweisen
- Nachweis von Zeitschriftenaufätzen, Büchern, Reports, Sammelwerksbeiträgen (ab Erscheinungsjahr), Dissertationen (ab 1968), audiovisuelle Medien (ab 1932)
- Publikationen (englisch, deutsch) von Autoren aus deutschsprachigen Ländern
- Aktualisierung: Monatlich (DlMDI), Vierteljährlich (GBI, Ovid Technologies/SilverPlatter)

Bei Online-Präsentationen bei Bedarf auf die interaktive Schaltfläche neben dem Titel klicken, damit wird automatisch die Internetverbindung zur entsprechenden Seite aufgebaut. Um aus dem Internet zum aktuellen Stand der Präsentation zurückzukehren, schließen Sie bitte die Internetansicht mit einem Klick auf das Kreuz oben rechts.

PSYINDEX

eine in deutscher und englischer Sprache absuchbare Datenbank mit Kurzreferaten. Sie weist weitgehend vollständig die psychologische Literatur (Zeitschriftenaufsätze, Bücher, Sammelwerksbeiträge, Reports, Dissertationen) und einschlägige audiovisuelle Medien nach. Dokumentiert werden deutsch- und englischsprachige Publikationen von Autoren aus Deutschland, Österreich und der Schweiz. An audiovisuellen Medien werden nur solche erfasst, die gekauft oder ausgeliehen werden können.

Fachgebiete: Gesamte Psychologie, einschließlich psychologisch relevanter Aspekte aus Nachbardisziplinen wie Psychiatrie, Soziologie, Erziehung, Philosophie, Sport, Kriminologie, Linguistik und angewandter Ökonomie.

Umfang: Publikationen ab Erscheinungsjahr 1977 (Dissertationen ab 1968), audiovisuelle Medien ab 1932.

Die Literatur- und Medien-Datenbank PSYDEX wurde 2002 um 8.119 Nachweise deutsch- und englischsprachiger Publikationen und AV-Medien ergänzt. Die Aktualisierung durch das ZPID erfolgt monatlich. PSYDEX umfasste Anfang 2003 über 170.000 Nachweise.

Stand 07.02.03: 170.128

Erschließung: Bibliographische Angaben, Kurzreferate, Schlagworte

Aktualisierung: Monatlich (DIMDI), vierteljährlich (G BI, Ovid Technologies/SilverPlatter).

Zusatzinformation Reports:

Was sind Reports?

Reports (Forschungsberichte, Bericht, Technischer Bericht, Final Report, Progress Report, Technical Note u. ä.) sind Dokumente, die formell die Ergebnisse oder Fortschritte eines Forschungs- und Entwicklungsvorhabens beschreiben. Reports werden unmittelbar nach oder während eines Forschungsprozesses geschrieben und enthalten daher Angaben über den neuesten Stand der Forschung, oft ausführliche Versuchsbeschreibungen und verschweigen auch negative verlaufene Untersuchungen nicht.

Reports gehören zur sogenannten Grauen Literatur, sind also in der Regel nicht über den Buchhandel zu beziehen und müssen direkt bei den herausgebenden Institutionen bestellt werden.

Abb. 2: Zusammenspiel von Folie und Notizseite am Beispiel einer Informationseinheit zur Datenbank PSYINDEX

Anhand dieser Parameter kann den Nutzern vermittelt werden, welche Informationen zur kritischen und adäquaten Beurteilung von Wissen herangezogen werden können und müssen. Da dieser elementare Bestandteil von Informations-

kompetenz bei einem Grossteil der Nutzer defizitär ist, liegt hier ein besonderer, ausbaufähiger Ansatzpunkt.

„In der so genannten Informationsgesellschaft ist es insbesondere in Forschung und Lehre unerlässlich, über aktuelle Informationen zu verfügen. Um diese Aktualität zu gewährleisten, ist es jedoch nicht ausreichend auf relevante Informationsquellen zugreifen zu können. Man muss auch in der Benutzung der Quellen geschult sein und die gefundene Information vernünftig be- und verwerten können. Den Studierenden soll nicht nur die Handhabung eines Werkzeugs beigebracht werden, sie sollen auch lernen, Informationen und ihren Produzenten kritisch gegenüberzutreten“ (Buoso, Kugler, Unterpertinger, 2002).

Bedienungskomfort

Aspekte des Bedienungskomforts sind relevante Variablen, um die Akzeptanz und Verwendung der angebotenen Lehr-Lernmodule zu erhöhen. Je angenehmer das Arbeiten mit den Materialien, desto größer ist ihre Akzeptanz – und damit die Verwendung. Daher wurde in der ZPID-Konzeption Wert auf hohen Bedienungskomfort gelegt.

Mit Hilfe von Hyperlinks in das Internet und einer Navigationsleiste auf den Folien werden den Nutzern entsprechende Benutzungsoptionen angeboten. Hyperlinks auf den Folien ermöglichen im Online-Modus einen schnellen und direkten Zugriff auf die entsprechenden Internetseiten. Dieser Zugriff und die damit verbundene Möglichkeit auf Fragen flexibel zu reagieren erhöht die Aufmerksamkeit und damit die Effektivität und Nachhaltigkeit der vermittelten Lehrinhalte.

Navigationsleisten verlinken zudem über interaktive Schaltflächen die Seiten innerhalb des Dokumentes. Sie können als schnelle Navigationshilfe durch die Präsentation genutzt werden. Bei der Verwendung von Overheadfolien dient die Navigationsleiste als Orientierungshilfe und Überblick. Eine Hinweisfolie im ersten Abschnitt der Präsentation verdeutlicht dem Anwender diese Strukturen (vgl. Abb. 3).

Wird nicht im Online-Modus präsentiert, sind zur Veranschaulichung aktuelle Screenshots in die Präsentation integriert. Diese können durch Angabe der Links im Notizfeld jederzeit ohne großen Aufwand vom Nutzer aktualisiert werden.

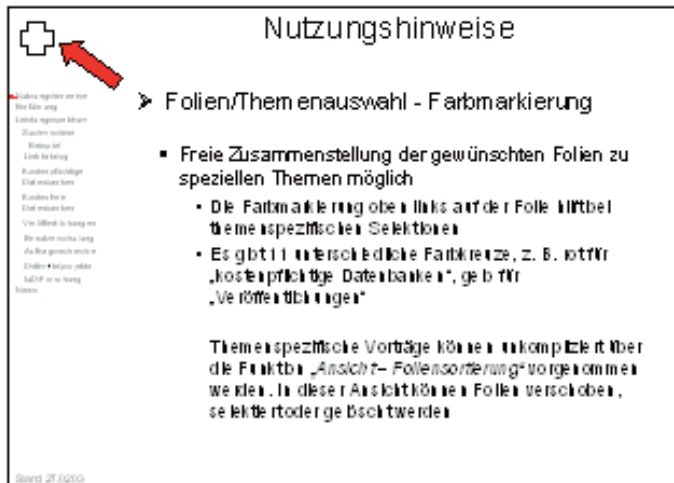


Abb. 3: Nutzungshinweise zu interaktiven Schaltflächen und Navigationsoptionen

Ausblick

Aktuell werden online zum einen Informations- und Präsentationsfolien zur Darstellung der verschiedenen Inhalte des Psychologie-Fachportals ZPID mit Fachdatenbanken und vielfältigen Recherchemöglichkeiten angeboten. Zum anderen helfen verschiedene Schulungsunterlagen, die Recherchekompetenz in den Fachdatenbanken zu optimieren.

Weitere spezifische Datenbankhilfen und interaktive Module sind in Planung. Durch Animation und interaktive Verlinkung soll die Informationskompetenz auf ansprechende Weise nachhaltig erhöht werden. Eine qualitative Nutzerbefragung soll dabei helfen, den Erfolg vorhandener Lehr- und Lernmodule zu messen und den individuellen oder institutsspezifischen Bedarf für geplante Projekte zu ermitteln.

Literatur

- Buoso, P., Kugler, U. & Unterpertinger, I. (2002). Freie Universität Bozen – Informationskompetenz. Verfügbar unter URL: <http://www.unibz.it/library/informationcompetence/?LanguageID=DE> (2003-04-15).
- Hapke, T. (2002). Informationskompetenz und studentisches Lernen im elektronischen Zeitalter Problemfelder und Praxisansätze in einer Technischen Universität, Vortrag auf dem 8. Kongress der IuK-Initiative der Wissenschaftli-

- chen Fachgesellschaften in Ulm 10. - 13. März 2002, Session 6: Informationskompetenz in Wissenschaft und Öffentlichkeit. Verfügbar unter URL: <http://www.tu-harburg.de/b/hapke/infolit/ulm-ik.htm> (2003-04-15).
- Homann, B. (2001). Schwierigkeiten und neue Ansätze der Benutzerschulung in Deutschland, 67th IFLA Council and General Conference, August 16-25, 2001. Heidelberg: Universitätsbibliothek Heidelberg.
- Klatt, R., Gavriilidis, K., Kleinsimlinghaus, K. & Feldmann, M. (2001a). Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung, Barrieren und Potenziale der innovativen Mediennutzung im Lernalltag der Hochschulen, Kurzfassung. Verfügbar unter URL: <http://www.stefi.de/download/kurzfass.pdf> (2003-04-15).
- Klatt, R., Gavriilidis, K., Kleinsimlinghaus, K. & Feldmann M. (2001b). Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung, Barrieren und Potenziale der innovativen Mediennutzung im Lernalltag der Hochschulen, Endbericht. Verfügbar unter URL: <http://www.stefi.de/download/bericht2.pdf> (2003-04-15).
- Schefczik, M. (2001). „Amerika, Du hast es besser?“ Erfahrungen zum Thema Information Literacy. Verfügbar unter URL: <http://www.uni-koblenz-landau.de/bibliothek/bielefeld.pdf> (2003-04-15).
- Zentrum für Psychologische Information und Dokumentation (2003). ZPID Präsentation. Verfügbar unter URL: <http://www.zpid.de/index.php?wahl=zpid&uwahl=presentation> (2003-04-15).

LIMES – A System for a Distributed Database Production in Mathematics

Olaf Ninnemann, Bernd Wegner

Zentralblatt MATH, FIZ Karlsruhe

Abstract

Zentralblatt MATH is the most comprehensive and traditional database providing information on publications in mathematics. With a variety of advantageous user-friendly features, which are associated with the search menu of the web version, it provides a large infrastructure for research and education in mathematics and its applications. Combined access to Zentralblatt MATH and related services are offered through the European Mathematical Information Service EMIS, which is a co-operation between several partners worldwide under the auspices of the European Mathematical Society EMS. Supporting the installation of a more European platform for Zentralblatt MATH, the project LIMES deals with the enhancement of Zentralblatt MATH to a European database in mathematics and with a prototype for a framework to distribute the editorial work to a European network. The aim of this article is to give a status report on LIMES and an idea of future actions for the implementation of the LIMES results in a European network.

0 Introduction

Databases with information on scientific literature emerged from the printed reviewing services with the rapid development of electronic devices for the publication of papers. Even before this development began the printed reviewing services were necessary to get an overview on the increasing production of scientific publications. But with the extended search facilities and the development of the web, providing convenient access by a variety of possibilities for links, these databases have become even more important as large infrastructures for research and education in science.

In order to support the role of the reviewing service Zentralblatt MATH as a large infrastructure, the project LIMES has been installed. It is a project that is guided by the EMS and funded within the Fifth Framework Programme (FP5) of the European Community. The acronym LIMES stands for „Large Infrastruc-

ture in Mathematics - Enhanced Services“. The project fits into the horizontal programme „Improving human research potential and the socio-economic knowledge base“ of the FP5, providing and improving access to research infrastructures. The duration of the project is from April 2000 to March 2004.

The partners of the LIMES-project are: FIZ - Fachinformationszentrum Karlsruhe (Germany), UJF - Cellule de Coordination Documentaire Nationale pour les Mathématiques (France), Eidetica (The Netherlands), SIBA - University of Lecce (Italy), DTV - Technical Knowledge Centre & Library of Denmark (Denmark), USDC - University of Santiago de Compostela (Spain), HMS - Hellenic Mathematical Society (Greece), TUB - Technical University of Berlin (Germany), and, as the supervising society, the EMS. The project co-ordinator is FIZ. The contracts for the funding of the individual projects were signed in March 2000. The work started on time on April 1, 2000.

The reviewing service Zentralblatt MATH provided the first reference database for mathematics, starting in 1978, as part of a network called STN. When mathematicians became more and more attracted by the Internet for their mutual communication and information, Zentralblatt also installed a web-based service. This service has been enhanced permanently by links to other electronic offers and facilities to use it for integrating information from Zentralblatt in primary publications. In principle, mathematicians and other researchers can access directly from their desktops everything needed from mathematics for their daily works. To improve the access facilities, even classical mathematical literature, which initially had only been printed and which remains of permanent interest in mathematics, has been digitised gradually through projects like ERAM (see [1] or [3]) and other digitisation projects like JSTOR and NUMDAM and made accessible on-line.

1 The database Zentralblatt MATH

The web-based service of Zentralblatt MATH is available via EMIS (see the URL <http://www.emis.de/ZMATH>). The major enhancement provided by this service consists in the indexing and the reports on the mathematical literature written by a large group of international experts. All reviewers work for Zentralblatt MATH more or less as volunteers. Further important aspects of quality are: comprehensiveness, precision of data, easy search menus, convenient linking with full texts and document delivery systems, etc. Mathematics as a science, where the truths detected by mathematicians do not lose their validity with the passage of time needs such precision in order to maintain accurate records of the achievements in the past and to obtain reliable information on what is new.

EMIS provides one link to the WWW-access, and, for the worldwide distribution, additional access has been arranged through 11 mirrors of Zentralblatt

MATH all over the world. To serve different mathematical communities, the partners and the LIMES team installed multilingual search interfaces. Most recently, starting with the ICM 2002 in Beijing, an interface in Chinese had been developed.

Searches can be made using the following list of fields: authors, titles, classification (MSC 2000), basic index, source, language, year of publication, etc. A search can be formulated in logical combinations of these terms. For the search, a graphic menu is available. The information is given in the AMSTEX source code, but several choices for a convenient formula display are available. Links to the full text of the corresponding article have been installed, if this is available electronically. Other options to get the full text of such an article consist of central document delivery services. Buttons to connect to such services and to see if the corresponding article is available there are installed in the menu for the search.

2 Related projects for accessing literature in mathematics and applications

There are two projects with homepages hosted by EMIS which are designed to provide services related to Zentralblatt MATH. One of them, called EULER, was funded by the European Union and is established as a service supported by an ever growing consortium. The other one is the Jahrbuch-Project ERAM, funded by the Deutsche Forschungsgemeinschaft.

The goal of the EULER project was to develop a prototype of an Internet service which integrates some of the most relevant publication-related electronic resources in the field of mathematics. This should provide „one-stop-shopping“ in heterogeneous resources for the user, for instance links to OPACs, databases, pre-print servers, electronic journals and other WWW-catalogues. More details on the project and the resulting service can be found in [2], [4] and in another presentation in this volume.

The aim of the Jahrbuch-Project, which officially is called ERAM (Electronic Research Archive in Mathematics), is to capture the Jahrbuch über die Fortschritte der Mathematik as a classical bibliographic service in mathematics in a database and to use this activity to select important publications from the Jahrbuch period for digitisation and storage in a digital archive. Longevity is typical for research achievements in mathematics. Hence to improve the availability of the classical publications in mathematics and to enable quick information retrieval of these, electronic literature information services and digital archives of the complete texts are needed as important tools for mathematical research in the future. This is the reason why so many efforts are invested in projects like ERAM.

The Jahrbuch-database will not be just a copy of the printed bibliography. It will contain a lot of enhancements.

The digital archive built up in connection with the database will be linked to the database and provide all facilities associated with current digitisation projects. More detailed reports, which also include the archiving activities, have been given in [3] and [1]. All data will have been captured by the end of 2003, and more than 30% has been enhanced already with the help of volunteers. The integration of the Jahrbuch data into the Zentralblatt database was arranged for in April 2003.

3 The LIMES project

The general objective of LIMES is to upgrade the database Zentralblatt MATH into a European based world-class database for mathematics and its applications by a process of technical improvements and wide Europeanisation. Upgrading the existing database, improving the present system and developing a new, distributed system both for the input and output of the data are necessary to allow Zentralblatt MATH to use the latest developments and to anticipate future developments of electronic technologies. This will make Zentralblatt MATH a world reference database, offering full coverage of the mathematics literature worldwide, including bibliographic data, reviews and/or abstracts, indexing, classification, excellent search facilities and links to offers of articles, with a European basis.

Improvements will be made in three areas:

1. Improvement of content and retrieval facilities through sophisticated further development of the current data sets and retrieval programs.
2. Broader and improved access to the database via national access nodes and new data distribution methods; in particular, improvement of access for isolated universities in regions with economic difficulties and in associated states of Central and Eastern Europe, where a mathematical tradition of excellence is under economic threat; stimulation of usage for all kinds of research as well as usage by funding organisations before making decisions for an initial support from two national test sides.
3. Improved coverage and evaluation of research literature via nationally distributed editorial units, development of technologies for efficient database production (exemplified by two further European member states).

The LIMES project sets out to achieve these goals by stimulation of usage through national access nodes and centres for dissemination including development of adequate licensing models. Links to offers of complete documents will

be an important addition. Better coverage and more precise evaluation of the literature through improved update procedures (for the data-gathering, but also for data distribution to access points) via a network of distributed Editorial Units will increase the quality of the database. Zentralblatt MATH will be extended to a European research infrastructure with distributed sites (Editorial Offices and Access Nodes) in the member states. The EMS, which represents the scientific communities, has promoted these ideas among the national societies and individual members.

4 The role of the LIMES participants

The work for LIMES is organised in several work-packages. The participants have been chosen such that each package is taken care of by one or several participants according to their special expertise.

FIZ operates the central editorial office for Zentralblatt MATH and is engaged in all work-packages of the project. The French partner, Cellule MathDoc (as a part of UJF), released edbm/w3, the first module of the European Database Manager for Mathematics software. In LIMES, UJF is in charge of the extensive development of this software, for example the integration of new tools for search, identification, display, and update, the development and integration of improved access control management, and it helps national access nodes with the development and installation of improved update procedures. Eidetica will mainly develop new models and tools for identification of author and serial names (two of the most requested features from the user community) based on data-mining techniques, and they will also develop other improvements to the retrieval process. An important task for TUB is the development of editorial tools.

The role of DTV is two-fold: 1. Setting up and running an editorial unit. The unit is responsible for covering a number of journals and other sources, which have been defined and agreed upon by DTU/MATH/DTV and the co-ordinator. The unit engages in experimenting with modern, electronic reviewing methods. 2. Establishing a raw data system for the editorial units. Once agreements have been reached with the publishers, this data will be made available for LIMES in two ways: a) for end-user current awareness purposes through a special LIMES gateway added to the Current Awareness System of the EMS and to the Zentralblatt MATH database, b) for the storage of preliminary metadata by the editorial units through the export of raw cataloguing data to the Zentralblatt editorial system. SIBA is the second partner to develop a streamlined system for the electronic handling of the reviewing workflow. They are supposed to design this in tight cooperation with DTV and the editorial office of Zentralblatt in Berlin. Nevertheless, having alternatives and additions to the solution developed by

DTV available, will be useful and enable the partners in the extended network to make their choice according to their special situation.

The role of USDC is to propagate the use of the database in Spain and to facilitate the access to it. They have installed a national site server and regularly update the database and new software releases. They have set up a task group for stimulating the usage of the database by Spanish universities, research centres, and libraries. A network structure for information in mathematics is in development, integrating Zentralblatt MATH with national information offers. This work is undertaken in co-operation mainly with CESGA (Centro de Supercomputación de Galicia) and RSME (Real Sociedad Matemática Española). HMS applies a methodology similar to that of the partner USDC in a different environment. They are quite successful in establishing a Greek access consortium. The EMS plays two roles in the project: It represents the end-user and gives a pan-European framework to the project as a whole. In both roles, it is part of the Project Management.

5 Some first achievements

The co-operation between the editorial centre of Zentralblatt in Berlin and its external partners has been restricted by the current input and administration procedures which were designed for a totally centralized workflow. Thus, the external editorial partners need continued, extensive support from the collaborators in Berlin in order to prevent bottlenecks in the workflow. One of these bottlenecks was the reviewer database, which provides all information needed for editors to assign papers to reviewers. Access from outside of the Zentralblatt office in Berlin was previously impossible. This system has been redesigned and installed on a platform which now allows for controlled access from outside. Clearly, such a database cannot be made open for public access, but the authorized partners can check the interests and the workload for every reviewer, they can register papers to be assigned to the reviewer, and they can make changes in the reviewer's data in accordance with the centre in Berlin.

A long-standing problem was to increase the flexibility of the input system for the database so as to be able to handle external input and to insert electronic submissions more systematically. The current system has a highly sophisticated structure with the aim to serve for a variety of different purposes. The input for the mathematics database is one of them; however, this is not the highest priority. Hence LIMES could not rely on a modification of the current system and had to design a new one that is specifically related to Zentralblatt and suitable for distributed users. A prototype for the new input system is available now and ready for being tested in a first release.

Links from references of articles to their review in a reference database are becoming a more and more important facility in electronic publications. UJF has developed a look up tool which will help to produce these links automatically based on the electronic references of an article. The tool has a surprisingly high precision. Also, some first tools have been developed for author identification with relatively good reliability. To do the same for journals will be comparatively easy. The models for the electronic reviewing flow are making good progress.

References

- [1] Hans Becker, Bernd Wegner: ERAM - Digitisation of Classical Mathematical Publications, Proc. ECDL 2000, Lecture Notes in Computer Science 1923, 424-427 (2000).
- [2] Bernd Wegner: EULER - a DC-based Integrated Access to Library Catalogues and Other Mathematics Information in the Web. Seventh International Conference „Crimea 2000" Libraries and Associations in the Transient World: New Technologies and New Forms of Co-operation. Conference Proceedings. Sudak, Autonomous Republic of Crimea, Ukraine, June 3-11, 2000, Volume 1: 264-267.
- [3] Bernd Wegner: ERAM - Digitalisation of Classical Mathematical Publications. Seventh International Conference „Crimea 2000" Libraries and Associations in the Transient World: New Technologies and New Forms of Co-operation. Conference Proceedings. Sudak, Autonomous Republic of Crimea, Ukraine, June 3-11, 2000, Volume 1: 268-272.
- [4] Bernd Wegner: EULER - a DC-based Integrated Access to Library Catalogues and Other Mathematics Information in the Web, ECDL 2000, Lecture Notes in Computer Science 1923, 461-466 (2000).

Contact

Olaf Ninnemann, Bernd Wegner
Zentralblatt MATH,
Franklinstr. 11
D - 10587 Berlin

Integration von digitalen Bibliotheken in Learning-Management-Systeme

Frank Oldenettel, Michael Malachinski

Oldenburger Forschungs- und Entwicklungsinstitut für
Informatik-Werkzeuge und –Systeme (OFFIS)

Zusammenfassung

In diesem Beitrag werden Ergebnisse des DFG-Forschungsprojektes LEBONED präsentiert, das die Integration digitaler Bibliotheken in web-basierte Lernmanagement-Systeme behandelt. Dazu wird zunächst erläutert, welche Modifikationen an einer Standard-Architektur von Lernmanagement-Systemen erforderlich sind, um eine solche Integration realisieren zu können. Einen bedeutenden Teil dieser Modifikationen stellt die so genannte LEBONED-Metadaten-Architektur dar. Diese beschreibt die Vorgehensweise hinsichtlich Metadaten und Dokumenten beim Importvorgang von einer digitalen Bibliothek in eine Lernumgebung sowie bei deren weiterer Verarbeitung. Die wesentlichen Komponenten dieser Architektur und deren Zusammenwirken werden detailliert in diesem Beitrag erläutert. Dazu zählen insbesondere die Aspekte der Anbindung digitaler Bibliotheken, Beschreibung von Dokumenten mittels geeigneter Metadaten, Fragmentierung von Dokumenten und Komposition von Fragmenten zu neuen Lernmaterialien.

Abstract

With this contribution we present results of the DFG funded research project LEBONED. This project deals with the integration of digital libraries into web based learning management systems. First, we explain how the architecture of a conventional learning management system has to be modified in order to enable the integration task. One important part of this modification is represented by the so-called LEBONED-Metadata-Architecture. It describes how to proceed if documents and metadata from a digital library shall be imported into the learning environment. The significant components of this architecture will be explained in detail with this contribution. In this context important aspects are connection of digital libraries to the learning environment, description of documents by metadata, fragmentation of monolithic documents, and composition of learning objects.

1 Einleitung

Mit der ständig wachsenden Ausdehnung des Internets erhalten immer mehr Menschen Zugang zu weltweit verteilten Informationsquellen, wodurch das Internet von einer permanent ansteigenden Zahl von Nutzern auch als ein Lehr-/Lernmedium wahrgenommen wird. Hochschulen gehen verstärkt dazu über, Online-Studiengänge anzubieten [1, 2]. [3] prognostiziert, dass virtuelle Universitäten den Bereich des e-Learning bis zum Jahre 2005 in signifikantem Maße beeinflussen werden. Die Durchführung von Lehrveranstaltungen erfolgt hier in unterschiedlicher Art und Weise. Entweder werden mit entsprechenden Telepräsentationstools Vorlesungen direkt auf den Rechner des Studierenden übertragen, oder es werden Lehr-/Lernmaterialien (Online-Kurse, Skripten, Foliensammlungen, Übungen, Musterlösungen etc.) im Rahmen einer web-basierenden Lehr-/Lernumgebung zur Verfügung gestellt. Kernstück einer solchen Lehr-/Lernumgebung bildet ein so genanntes *Learning Management System* (LMS), das als Web-Portal eine Reihe von Diensten wie z.B. die Verwaltung der Online-Kurse oder Prüfungsfunktionalitäten bereitstellt. Detailliertere Beschreibungen können [2] oder [4] entnommen werden.

[2] definiert eine Reihe von Anforderungen, die ein LMS erfüllen sollte. Grundlegend ist dabei die effiziente Verwaltung von elektronischen Lehrmaterialien innerhalb eines Content Repositorys. Diese Anforderung wird i.d.R. von den bereits existierenden Systemen (siehe dazu [5]) in unterschiedlichem Umfang erfüllt. Eine weitere wichtige Anforderung, die von [2] beschrieben, aber leider von keinem der heutigen LMS erfüllt wird, ist die Fähigkeit, externe Wissensressourcen zu integrieren. Das bedeutet, dass der Umfang an digitalen Lehrmaterialien, der Studierenden in einer Lehr-/Lernumgebung zur Verfügung steht, auf den Bestand des lokalen Content-Repositorys beschränkt ist.

Dem gegenüber steht eine Vielzahl von digitalen Bibliotheken, die über das Internet zugänglich sind. Diese Systeme verwalten digitale, oft multimedial aufbereitete Dokumente und verfügen über die dafür notwendigen Speicherungs- und Zugriffsmechanismen. Daher sind digitale Bibliotheken prädestiniert für die Verwaltung von elektronischen Lehrmaterialien.

Da digitale Bibliotheken bereits seit vielen Jahren Gegenstand zahlreicher FuE-Aktivitäten sind (siehe [6-8]), wurde lange bevor das allgemeine Interesse an virtuellen Lehr-/Lernumgebungen wuchs, eine Vielzahl digitaler Bibliotheken ohne konkreten Bezug zu Lernumgebungen entwickelt. Ein Großteil dieser Systeme hat sich etabliert und ist bis heute im Einsatz, wie z.B. die digitalen Bibliotheken von ACM¹ oder IEEE² oder z.B. das *eVerlage*-System [9], die *California Digital Library* [10], die *New Zealand Digital Library* [11, 12] oder die *Digitale Bibliothek NRW* [13], um nur einige zu nennen.

1 <http://www.acm.org/dl>

2 <http://www.computer.org>

Auch ohne Bezug zu einer Lernumgebung weisen viele digitale Bibliotheken Inhalte auf, die für Lernzwecke verwendet werden könnten. Allerdings sind die in diesen Systemen abgelegten Dokumente häufig nur eingeschränkt nutzbar, da sie sich i.d.R. auf die bloße Darstellung der Inhalte beschränken. Um mit diesen Dokumenten effektiv lernen zu können, müssen sie entsprechend aufbereitet werden. Es sollten z.B. Annotationen möglich sein, oder es sollte der Kenntnisstand des Lernenden berücksichtigt werden können. Auch wäre es hilfreich, Teile von Dokumenten mit Teilen anderer Dokumente zu kombinieren. Dabei müssen sie auch verschiedenen, heterogenen Quellen entstammen dürfen. Außerdem sind für solche neu konstruierten Dokumente geeignete Präsentationsmechanismen erforderlich, die ggf. auch Interaktionsmöglichkeiten bieten. Beim Umgang mit Dokumententeilen müssen verschiedene Dokumentenformate (Text, Grafik, Audio, Video, etc.) berücksichtigt werden.

Die Konstruktion neuer Dokumente aus Teilen verschiedener Materialien ist für Lernende und Lehrende gleichermaßen attraktiv. Der Lernende bekommt dadurch die Möglichkeit, sich Lernmaterial nach eigenen Gesichtspunkten zusammenzustellen. Aus Sicht des Lehrenden ergibt sich die Möglichkeit, Lehrmaterialien zu erstellen, um z.B. einen Online-Kurs vorzubereiten. Da laut [14] die Erstellung von elektronischen Lehr-/Lernmaterialien zukünftig zu einem der wichtigsten Prozesse im Hochschulumfeld gehören wird, ist es insbesondere aus Sicht der Lehrenden wünschenswert, Mechanismen zur Verfügung gestellt zu bekommen, mit denen auf bereits existierende Wissensressourcen, wie digitale Bibliotheken, zurückgegriffen werden kann.

Aus diesem Grund ist es nahe liegend, die für eine Integration von digitalen Bibliotheken notwendigen Technologien (Methoden, Infrastrukturen, Werkzeuge) zu entwickeln. Allerdings ergeben sich hierbei einige Probleme.

1.1 Probleme

Ein wesentliches Problem bei der Integration von digitalen Bibliotheken in LMS besteht, darin, dass für die Dokumente aus den Bibliotheken keine adäquaten Metadaten vorhanden sind, um im Content-Repository des LMS angemessen verwaltet werden zu können. Digitale Bibliotheken sind i.d.R. nur in der Lage, bibliographische Metadaten zu liefern, die entweder einem bekannten Standard wie z.B. *Dublin Core (DC)*³ [15], *Encoded Archival Description (EAD)*⁴ [16], *MARC21*⁵ [17] oder einem proprietären Format entsprechen. Diese bibliographischen Informationen sind für Lernzwecke unzureichend. Hier werden weiterführende Informationen, wie z.B. das Lernziel oder der Schwierigkeitsgrad eines Dokuments benötigt.

3 <http://www.dublincore.org/>

4 <http://www.loc.gov/ead/>

5 <http://www.loc.gov/marc/>

Metadatenformate, die diese Anforderungen erfüllen, sind in Standards wie *Learning Object Metadata (LOM)*⁶ [18], *Sharable Content Object Reference Model (SCORM)*⁷ [4], *Aviation Industry Computer Based Training Committee (AICC)*⁸ oder *Instructional Management Systems Project (IMS)*⁹ definiert. Allerdings ist die Unterstützung bibliographischer Meta-Informationen bei diesen Formaten eher schwach ausgeprägt. Für eine adäquate Verwaltung von Dokumenten aus einer digitalen Bibliothek innerhalb des Content-Repositorys eines LMS ist demnach ein Metadatenformat erforderlich, welches alle lernspezifischen Aspekte der zuletzt genannten Standards unterstützt und zusätzlich die wichtigsten bibliographischen Aspekte berücksichtigt. Ein solches Metadatenformat bildet dann die Grundlage für das Dokumentenmodell, das beliebige Dokumente, welche im LMS verwaltet werden sollen, sowohl inhaltlich als auch strukturell beschreibt. Auf Basis dieses Dokumentenmodells wird das Datenschema des Content-Repositorys definiert.

Da digitale Bibliotheken nur bibliographische Metadaten liefern können, fehlen viele lernspezifische Informationen und müssen während des Imports von Dokumenten hinzugefügt werden. Dies kann zum Teil automatisiert geschehen. Ein anderer Teil muss jedoch von einem Nutzer hinzugefügt werden. Für diesen Zweck ist eine entsprechende Methodik mit Werkzeugunterstützung erforderlich, um diesen Vorgang möglichst effizient, aber auch inhaltlich angemessen zu gestalten.

Eine weitere Form der Aufbereitung der importierten Dokumente während des Importvorgangs ist die Identifikation von Dokumentenstrukturen und die Fragmentierung der Dokumente anhand der identifizierten Strukturen. Da es - wie bereits eingangs erwähnt - sowohl für Lernende als auch für Lehrende hilfreich ist, Komponenten von Dokumenten in unterschiedlichen Kontexten (wieder)zuverwenden, ist es erforderlich, die importierten Dokumente in Form von einzelnen Komponenten, assoziiert mit entsprechenden Metadaten, als Lernobjekte im Content-Repository abzulegen. Allerdings haben die meisten der von digitalen Bibliotheken gelieferten Dokumente einen sehr monolithischen Charakter. Das bedeutet, dass z.B. ein ganzes Buch in Form einer einzigen physikalischen Ressource (Datei) geliefert wird. Um auf einzelne Teilkomponenten zugreifen zu können, müssen die internen logischen Strukturen identifiziert und die Dokumente entsprechend fragmentiert werden.

Ein weiteres Problem besteht im Fehlen geeigneter Schnittstellen zwischen LMS und den zu integrierenden digitalen Bibliotheken. Die Implementierung einer Mediator-Wrapper-Architektur (siehe [19, 20],) stellt eine gangbare Lö-

6 <http://grouper.ieee.org/groups/ltsc/wg12/>

7 <http://www.adlnet.org/index.cfm>

8 <http://www.aicc.org>

9 <http://www.imsproject.org>

sung dar. Im Falle der Integration einer digitalen Bibliothek in eine Lernumgebung ergeben sich jedoch noch zusätzliche Probleme:

- *Absetzen von Rechercheanfragen von der Lernplattform an die digitale Bibliothek:* Falls das LMS nicht über eine adäquate Recherchefunktionalität verfügt, kann diese auch nicht auf die zu entwickelnde Schnittstellenkomponente abgebildet werden. In diesem Fall ist eine Erweiterung der Recherchefunktionalität auf Seiten der Lernplattform notwendig.
- *Auswahl und Zugriff auf Rechercheergebnisse:* Im Anschluss an eine Dokumentenrecherche muss es durch geeignete Mechanismen möglich sein, die gewünschten Dokumente aus dem Rechercheergebnis auszuwählen und in den Dokumentenbestand der Lernplattform zu integrieren. Auch hier ist eine Erweiterung auf Seiten der Lernumgebung erforderlich.

Um für die in diesem Abschnitt beschriebenen Probleme Lösungsansätze zu entwickeln, wurde das Projekt LEBONED initiiert, das im folgenden kurz beschrieben werden soll.

1.2 Das LEBONED-Projekt

Das Projekt **LEBONED (Learning Environment Based on Non Educational Digital Libraries)** wird im von der *Deutschen Forschungsgemeinschaft (DFG)* im Rahmen des Schwerpunktprogramms *Verteilte Verarbeitung und Vermittlung digitaler Dokumente (V3D2)* gefördert. Der Projektstart war im April 2002 und das Projektende ist für März 2004 vorgesehen.

Das Hauptziel von LEBONED besteht darin, eine Methodik für die Integration existierender digitaler Bibliotheken in Lernumgebungen zu entwickeln. Des Weiteren soll eine Infrastruktur zur Implementierung einer für diese Zwecke geeigneten Lernumgebung entwickelt werden. Diese soll in Form einer Mediator-Wrapper-Architektur realisiert werden. Existierende digitale Bibliotheken werden dabei von sog. Wrappern gekapselt und an die Lernumgebung angebunden. Zentrale Komponenten der Lernumgebung stellen Recherchemechanismen und eine Dokumentenverwaltung dar, deren Grundlage ein zu entwickelndes Dokumentenmodell bildet. Weitere Ziele dieses Vorhabens sind unter anderem:

- *Entwicklung einer komponentenbasierten Rahmenarchitektur für eine integrative Lernumgebung:* Diese Rahmenarchitektur soll die Voraussetzungen schaffen, die notwendig sind, um externe Systeme an ein LMS anzubinden und von dort bezogene Dokumente zu verwalten. Hierzu gehört zunächst die Umsetzung eines geeigneten Dokumentenmodells. Dies ist ein Metadatenformat, welches als Grundlage für die Dokumentenverwaltung dient. Des Weiteren sind Recherche- und Zugriffsmechanismen zu konzipieren, die es ermöglichen sowohl im lokalen Content-Repository des LMS als auch in den

externen digitalen Bibliotheken zu recherchieren und auf Dokumente zuzugreifen.

- *Entwicklung von Werkzeugen zur Aufbereitung von Dokumenten:* Hierunter ist sowohl die Ergänzung von fehlenden Metadaten zu verstehen, als auch die Fragmentierung von monolithischen Dokumenten zu Lernobjekten, sowie deren Neukomposition mit anderen Lernobjekten.
- *Entwicklung eines Vorgehensmodells zur Integration externer digitaler Bibliotheken:* Zusätzlich wird zum Vorgehensmodell ein Leitfaden erstellt, der Hilfestellung beim Integrationsprozess liefert.

Im Rahmen des LEBONED-Projekts wird das Hauptaugenmerk auf die technischen Aspekte gelegt, die bei der Integration von bereits existierenden digitalen Bibliotheken in eine Lernumgebung zu berücksichtigen sind. Als allgemeineres Projektziel wird angestrebt, eine Methodik und Software-Komponenten mit Referenzcharakter zu entwickeln.

Im nachfolgenden Abschnitt soll zunächst auf die grundlegende Struktur eines LMS eingegangen werden. Dort wird anhand einer idealtypischen Sicht dargestellt, welche Komponenten (Dienste) ein konventionelles LMS i.d.R. bietet und um welche Elemente ein solches System erweitert werden muss, um die Integration von digitalen Bibliotheken zu ermöglichen. Der dritte Abschnitt befasst sich dann mit einer konkreteren Betrachtung eines Teilaspektes einer integrativen LMS-Architektur, dem Import von Dokumenten und den dazugehörigen Metadaten.

2 Modell eines integrativen Learning-Management-Systems

Eine Betrachtung verschiedener LMS (siehe dazu [5]) zeigt, dass diese Systeme zwar sehr unterschiedlich sind, aber auch viele Gemeinsamkeiten aufweisen. Diese Gemeinsamkeiten spiegeln sich vor allem in den von den Systemen bereitgestellten Diensten wider. Dagegen sind die Unterschiede vor allem in den technischen Realisierungen bzw. Systemarchitekturen zu finden. Daher ist es nicht sinnvoll, den Versuch der Konstruktion einer generischen Architektur für Lernmanagement-Systeme zu unternehmen, die für alle Ausprägungen dieser Systemgattung Gültigkeit besitzt. Jedoch ist es aufgrund der Gemeinsamkeiten bzgl. der zur Verfügung gestellten Dienste durchaus möglich ein generisches Modell für Lernmanagement-Systeme anzugeben. Ein solches Modell bzw. die dort vorgesehenen Dienste werden z.B. in [4] beschrieben. In Abbildung 1 ist eine idealtypische Sicht eines Modells dargestellt, welches um einige zusätzliche Dienste erweitert wurde, die zur Integration von digitalen Bibliotheken erforderlich sind. Bereits im konventionellen Modell beschriebene Dienste sind in

Abbildung 1 weiß dargestellt, während die neu integrierten Dienste grau dargestellt sind. Die Mischfarbe des *Content-Management-Service* deutet an, dass dieser Dienst zwar bereits im konventionellen Design vorhanden war, im integrativen Design jedoch in modifizierter Form vorliegen muss. Bei den schraffiert abgebildeten Elementen handelt es sich um externe Komponenten, die nicht Bestandteil des LMS sind. Es folgt eine detaillierte Beschreibung der einzelnen Dienste.

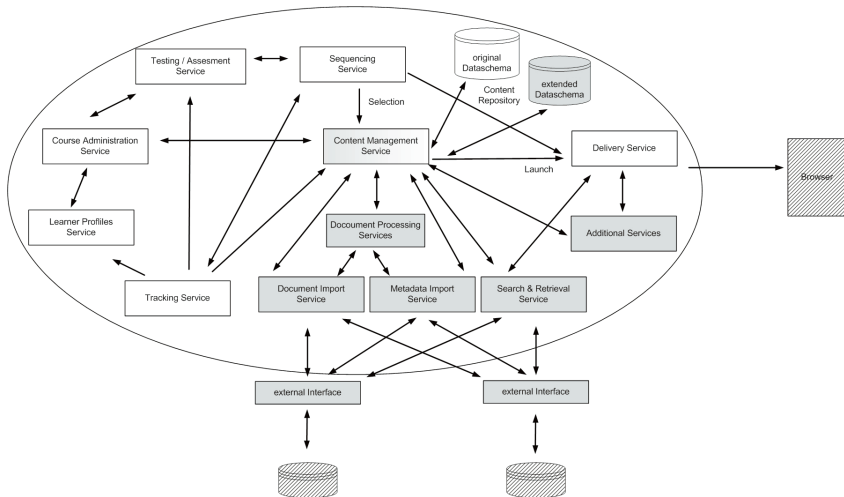


Abbildung 1: Generisches Modell für ein integratives LMS

Extended Dataschema: Da konventionelle LMS häufig keine adäquaten Speichermöglichkeiten für importierte Dokumente vorsehen oder sogar keine eigene Dokumentenverwaltung bieten [5], sondern sich vielmehr auf die Verwaltung von Kursen beschränken, dient das *Extended Dataschema* als Ergänzung zum originalen LMS-Datenschema der Verwaltung von importierten Dokumenten.

Content-Management-Service: Der *Content-Management-Service* ist der zentrale Dienst, über den alle Zugriffe auf das Content-Repository abgewickelt werden. Im Gegensatz zum konventionellen Modell muss dieser Dienst im integrativen Modell einen erweiterten Funktionsumfang aufweisen. Die bereitgestellten Recherche, Zugriffs- und Verwaltungsmechanismen müssen auf das *extended Dataschema* ausgeweitet werden. Der Content-Management-Service bietet eine einheitliche Sicht darauf, so dass sich beide Schemata als einheitliches Content-Repository darstellen.

Search- & Retrieval-Service: Dieser Dienst stellt die notwendige Funktionalität bereit, um nach elektronischen Lehrmaterialien recherchieren zu können. Dabei werden Suchanfragen in einem internen Anfrageformat formuliert und gleichermaßen an das lokale Content-Repository und an die angeschlossenen externen digitalen Bibliotheken gestellt. Die von den verschiedenen Repositories zurückgelieferten Anfrageergebnisse werden wieder entgegengenommen und für die weitere Verwendung aufbereitet.

Document-Import-Service: Dieser Dienst wickelt ggf. im Anschluss an eine Dokumentenrecherche den Import der gewünschten Dokumente ab. Aus der abstrakten Sicht dieses Modells ist die Funktionalität dieses Dienstes wenig komplex. Sie beschränkt sich auf das Abrufen der Dokumenten von den digitalen Bibliotheken und dem Einfügen in das lokale Content-Repository.

Metadata-Import-Service: Da zur Verwaltung innerhalb des Content-Repositorys des LMS entsprechende Metadaten benötigt werden, liegt es nahe sich der Metadaten zu bedienen, die bereits von den digitalen Bibliotheken bereitgestellt werden. Diese Aufgabe wird vom *Metadata-Import-Service* geleistet. Der Dienst ruft für jedes zu importierende Dokument die dazugehörigen Metadaten ab.

Document-Processing-Services: Unter der Bezeichnung *Document-Processing-Service* werden eine Reihe von Diensten zusammengefasst, die der Aufbereitung von importierten Dokumenten dienen. Dabei ist in erster Linie das Ergänzen von fehlenden Metadaten zu nennen, denn bevor ein importiertes Dokument als vollwertiges Lernobjekt innerhalb des Content-Repositorys recherchierbar ist, muss die dazu gehörige Metadatenbeschreibung vervollständigt werden. Ein weiterer Dienst sollte die Möglichkeit bieten, große monolithische Lernobjekte, wie z.B. ganze Bücher, in kleinere Lernobjekte (z.B. Kapitel) zu zerlegen, um diese gezielt recherchierbar und zugreifbar zu machen und um sie in unterschiedlichen Kontexten wieder verwenden zu können. Eine solche Wiederverwendung ist beispielsweise die Neukomposition verschiedener einfacher Lernobjekte zu neuen komplexeren Dokumenten. Auf diese Weise lassen sich Lehrmaterialien individuell für bestimmte Lernzwecke gestalten. Ein Dienst, der dieses ermöglicht, muss ebenfalls unter den *Document-Processing-Services* vorhanden sein. Da alle diese Dienste nicht vollständig automatisierbar sind, ist hier eine Interaktion mit dem Nutzer erforderlich. Daher sollten sie in Form von entsprechenden Werkzeugen angeboten werden.

External Interface: Die eigentliche Anbindung der digitalen Bibliothek erfolgt über die *external Interfaces*. Diese Komponenten kapseln die spezifischen Zugangsschnittstellen der digitalen Bibliotheken und stellen dem LMS somit einheitliche Zugriffsmöglichkeiten zur Verfügung. Die wichtigsten Aufgaben bestehen im Umsetzen von Rechercheanfragen in das spezifische Format der digitalen Bibliothek und dem Zurückliefern der Anfrageergebnisse.

Additional Services: Unter dem Begriff *Additional Services* sollen alle Dienste zusammengefasst werden, die für die Integration von digitalen Bibliotheken in LMS nicht zwingend erforderlich, jedoch in bestimmten Fällen von Bedeutung sind. Diese Dienste umfassen z.B. Funktionalitäten zur Nutzerauthentifizierung oder Bestellung und Zahlung von kostenpflichtigen Materialien.

Wie bereits erwähnt, handelt es sich bei dem hier vorgestellten Modell um eine abstrakte Darstellung, die beschreibt, welche Dienste für ein integratives Lernmanagement-System grundsätzlich von Bedeutung sind. Im folgenden soll der Aspekt des Dokumenten- und Metadatenimports herausgegriffen und näher betrachtet werden.

3 Die LEBONED-Metadatenarchitektur

Die LEBONED Metadaten-Architektur ist kein eigenständiges Modell, sondern bildet einen Teilbereich des zuvor beschriebenen Gesamtmodells. Hier werden alle Komponenten und Vorgänge beschrieben, die am Import von Dokumenten und den dazu gehörigen Metadaten aus externen digitalen Bibliotheken beteiligt sind.

Die Search-Retrieval-Komponente ist nicht Bestandteil der hier beschriebenen Metadaten-Architektur, denn der durch dieses Modell beschriebene Prozess beginnt an dem Punkt, an dem ein Recherchevorgang abgeschlossen ist und die gewünschten Dokumente bereits ausgewählt sind. Abbildung 2 zeigt eine Darstellung der LEBONED Metadaten-Architektur.

Der Importvorgang des Modells beginnt bei den *Wrappern*, (external Interfaces) die sowohl die gewünschten Dokumente als auch die dazugehörigen Metadaten von den digitalen Bibliotheken beziehen. Dass Wrapper je nach Zugangsmöglichkeiten zu den digitalen Bibliotheken unterschiedlich realisiert sein können (web-basiert, webservice-basiert, protokoll-basiert oder API-basiert), spielt in dieser Darstellung keine Rolle. Unabhängig von der Realisierung hat jeder Wrapper die gleichen Aufgaben.

vorliegenden Metadaten in das Zielformat *lxSCORM*, auf dem das Dokumentenmodell des Content-Repositorys basiert. *lxSCORM* ist eine auf dem SCORM-Standard basierende Erweiterung, die in Abschnitt 3.1 noch etwas genauer vorgestellt wird.

Nach der Konvertierung der von den digitalen Bibliotheken bezogenen Metadaten liegen diese zwar nun im endgültigen Format vor. Sie können dennoch nicht gemeinsam mit den dazugehörigen Dokumenten in der Dokumentenverwaltung des Content-Repositorys abgelegt werden, da sie noch unvollständig sind. Digitale Bibliotheken liefern größtenteils nur bibliographische Metadaten, jedoch keine lernspezifischen Informationen. Diese Daten müssen nachträglich hinzugefügt werden. Da diese Aufgabe nicht automatisiert ablaufen kann, wird an dieser Stelle das Tool *lexSCORE* eingesetzt, mit dessen Hilfe ein Benutzer die fehlenden Metadaten effizient hinzufügen kann. Das Tool lädt neben den Metadatenbeschreibungen, die es von der Metadaten-Import-Komponente erhält, das dazu gehörige Dokument und ermöglicht dem Nutzer durch entsprechende Darstellung des Dokuments oder einzelner Teile davon, fehlende Metainformationen einfach und bequem hinzuzufügen.

Nach der Bearbeitung der Metadatenbeschreibung mit Hilfe von *lexSCORE* liegen die Metadaten vollständig im Zielformat *lxSCORM* vor und können gemeinsam mit den dazugehörigen Dokumenten in der Dokumentenverwaltung abgelegt werden. Hier wird zunächst jedes Dokument in seiner ursprünglichen Form als monolithisches Lernobjekt im Content-Repository gespeichert, während die damit assoziierten Metadaten im Metadata-Repository abgelegt werden. Von nun an stehen diese Lernobjekte für weitere Bearbeitungen, die jederzeit durchgeführt werden können, zur Verfügung. Damit ist der Pfad, den Dokumente und Metadaten ausgehend von den digitalen Bibliotheken zur Dokumentenverwaltung des LMS nehmen, abgeschlossen.

Weitere Bearbeitungen von Lernobjekten, die bereits in der Dokumentenverwaltung des LMS vorliegen, werden in Abb. 2 durch weitere Pfade in Form von Schleifen dargestellt, die an der Dokumentenverwaltung beginnen und dort auch wieder enden. Mit Hilfe des Tools *Learning-Object-Creator* können „größere“ Lernobjekte in mehrere „kleine“ fragmentiert werden. Dazu werden monolithische Dokumente bzw. Lernobjekte oder bereits aus einer früheren Fragmentierung hervor gegangene „größere“ Lernobjekte aus der Dokumentenverwaltung in den *Learning-Object-Creator* geladen. Das Tool unterstützt den Nutzer bei der Identifikation von internen logischen Dokumentenstrukturen und ermöglicht eine physikalische Fragmentierung des Lernobjekts anhand dieser Strukturen. Das Tool *Learning-Object-Composer* kann dazu verwendet werden, bereits existierende Lernobjekte, die ursprünglich aus verschiedenen Dokumenten hervorgegangen sind, in neuen Kontexten zusammenzufügen und somit neue Lehrmaterialien zu gestalten, die individuell auf spezifische Bedürfnisse zugeschnitten werden können. Sowohl nach einer Fragmentierung eines mono-

lithischen Dokuments mit dem Learning-Object-Creator als auch nach einem Neuarrangement von Lernobjekten mit dem Learning-Object-Composer liegen für die erzeugten Fragmente bzw. Kompositionen keine vollständigen Metadatenbeschreibungen vor. Die zunächst nur mit den Ausgangsdokumenten assoziierten Metadaten, sind nicht ohne weiteres auf die Resultate anwendbar. Daher werden die unvollständigen Metadatenbeschreibungen wieder mit dem Tool *lexSCORE* vervollständigt, bevor die erzeugten Dokumente bzw. Lernobjekte in der Dokumentenverwaltung abgelegt werden.

Wie bereits erwähnt wird eine solche Metadatenbeschreibung in dem auf SCORM basierenden Format *lxSCORM* erzeugt, das im Folgenden kurz vorgestellt werden soll.

3.1 Metadatenformat *lxSCORM*

Der SCORM-Standard enthält unter anderem eine auf LOM basierende Metadatenpezifikation, die im wesentlichen alle Anforderungen an den Umgang mit Lernobjekten innerhalb einer konventionellen Lehr-/Lernumgebung erfüllt. Im Falle der Integration von digitalen Bibliotheken ergeben sich jedoch einige Anforderungen, die von diesem Standard nicht erfüllt werden können. Daher sind einige Erweiterungen an der Metadatenpezifikation erforderlich und dies führte zu der Entwicklung von *lxSCORM* (*leboned extended SCORM*). Hierbei, wurde der Abwärtskompatibilität zum ursprünglichen SCORM-Standard oberste Priorität beigemessen. Aus diesem Grund weist *lxSCORM* ausschließlich Erweiterungen in Form von zusätzlichen Elementen auf. An bereits existierenden Elementen wurden keinerlei Veränderungen vorgenommen. Auf diese Weise können auch andere SCORM-konforme Anwendungen *lxSCORM*-Beschreibungen verarbeiten, da die zusätzlichen Elemente einfach ignoriert werden. Hätte *lxSCORM* auch Modifikationen an bereits existierenden Elementen vorgesehen, könnte dies bei anderen Anwendungen zu Fehlinterpretation der Inhalte führen und eine Kompatibilität wäre nicht mehr gegeben.

Die Defizite von SCORM umfassen sowohl einige technische als auch einige bibliographische Aspekte. So ist es z.B. aus technischer Sicht mit einer SCORM-Beschreibung nicht ohne weiteres möglich die interne logische Struktur (Kapitel, Abschnitte, Szenen) eines Lernobjekts abzubilden. Zwar ist es möglich mehrere SCORM-Objekte zu erzeugen, die jeweils mit einem anderen Bereich der internen Struktur assoziiert sind, allerdings kann mit dem SCORM-Element 4.3 *Location* (siehe [21]) nur eine Referenz auf die ganzheitliche physikalische Ressource (Datei) angegeben werden. Mit der *lxSCORM*-Erweiterung 4.8 *Extended Location* und den Subelementen 4.8.1. *Begin* und 4.8.2 *End* (siehe Abbildung 3) ist es möglich, ein *lxSCORM*-Objekt auf Teilbereiche innerhalb einer physikalischen Ressource zu referenzieren. Weitere Defizite finden sich bei SCORM z.B. im bibliographischen Bereich, die

mit den lxSCORM-Erweiterungen 1.11 Bibliographic Extension, 1.12 Pages und 2.4 Periodicals behoben wurden.

Bei den hier vorgestellten Erweiterungen handelt es sich lediglich um einige Beispiele von lxSCORM-Elementen. Weitere Beispiele mit detaillierten Erläuterungen lassen sich [22] entnehmen.

Nr.	Name	Explanation	Multiplicity	Data Type
4.8	Extended Location	[...] (see original specification)		
4.8.1	Begin	Indicates the beginning of the part of the physical resource to which this metadata instance is related. Value type depends on item 4.1. Format. If item 4.1 indicates an continous format (audio, video) this item describes a time stamp If item 4.1 indicates an textformat (HTML, PDF, PostScript, RTF, etc.) this item describes a character position If item 4.1 indicates an graphical format this item describes coordinates	0 or 1	String
4.8.1	End	Indicates the end of the part of the physical resource to which this metadata instance is related. Value type depends on item 4.1. Format. If item 4.1 indicates an continous format (audio, video) this item describes a time stamp If item 4.1 indicates an textformat (HTML, PDF, PostScript, RTF, etc.) this item describes a character position If item 4.1 indicates an graphical format this item describes coordinates	0 or 1	String

Abbildung 3: Spezifikation des lxSCORM-Elements 4.8 Extended Location mit Subelementen

4 Verwandte Arbeiten

Als verwandte Arbeit soll insbesondere das von der Europäischen Union geförderte Projekt *ARIADNE* genannt werden, das sich unter anderem mit dem Zugriff auf verteilte Wissensquellen befasst. Die Integration existierender digitaler Bibliotheken in LMS wird dabei jedoch weniger explizit und detailliert berücks-

ichtigt als in LEBONED, deshalb kann das Projekt als Ergänzung hierzu betrachtet werden.

Weiterhin spielen für LEBONED die existierenden LMS eine wesentliche Rolle. Einen umfassenden Überblick hierüber gibt [5].

Ebenso sind für das Projekt natürlich alle existierenden digitalen Bibliotheken von Belang, deren Inhalte sinnvoll für Lernzwecke eingesetzt werden können. Dazu zählen beispielsweise die digitalen Bibliotheken der IEEE und der ACM, DSpace vom MIT oder die digitale Bibliothek eVerlage [9].

Im Bereich der verteilten digitalen Bibliotheken sei zunächst das Projekt *New Zealand Digital Library* (NZDL) der Universität Waikato genannt. Dort wurde das Softwarepaket *Greenstone* entwickelt, das den Aufbau und Betrieb eigener Kollektionen verteilter digitaler Bibliotheken ermöglicht. Eine weitere nennenswerte verteilte digitale Bibliothek ist die *California Digital Library* (CDL) der Universität von Kalifornien [10]. Ziele dieses Projekts sind insbesondere die nahtlose Integration verteilter Ressourcen und der Zugriff auf diese. In diesem Zusammenhang soll auch das *Stanford Digital Libraries Technologies Project* erwähnt werden, in dem die *InfoBus*-Technologie entwickelt wurde. Dabei handelt es sich um eine CORBA-basierte Infrastruktur zur Integration verteilter, heterogener Kollektionen digitaler Bibliotheken. Das Projekt *DAFFODIL* der Universität Dortmund beschäftigt sich mit der Entwicklung einer agentenbasierten Infrastruktur für föderierte digitale Bibliotheken [23].

Wichtig für LEBONED sind außerdem Arbeiten, die sich mit Metadaten im Kontext von e-Learning befassen. Das adaptive Hypermedia-System *Multibook* [24] beispielsweise wird zur Vermittlung von Multimedia-Technologien eingesetzt und generiert, basierend auf Metadaten, die Reihenfolge der Kursinhalte semi-automatisch. Das Projekt *CUBER* dagegen dient der Entwicklung eines Systems zur Suche nach Lernmaterialien an europäischen Universitäten. Dazu wurden Modifikationen an der LOM-Spezifikation vorgenommen [25]. Eine weitere in [25] beschriebene Arbeit behandelt Aspekte der Wiederverwendbarkeit und Adaptierbarkeit von multimedialen Lerninhalten in LMS. Auch hier wurden Erweiterungen des LOM-Standards durchgeführt.

Vom IMS wird neben Metadatenformaten auch ein Standard namens *IMS Digital Repository Interoperability* bereitgestellt, der festlegt, wie der Zugriff auf verteilte Repositories mit Lernobjekten erfolgen soll. Dieser Standard setzt jedoch zum einen das Vorhandensein wohldefinierter Lernobjekte in den Repositories voraus, zum anderen wird die Verwendung konkreter Technologien wie Z39.50, XQuery oder SOAP festgelegt, die von der Vielzahl digitaler Bibliotheken derzeit noch nicht unterstützt werden. Deshalb ist dieser Standard im Rahmen von LEBONED nicht anwendbar.

Mit Infrastrukturen und Werkzeugen für die Wiederverwendung von Lernobjekten beschäftigt sich das Projekt *Teachware on Demand* (www.teachware-on-demand.de). Ziel ist die automatische Generierung benutzer-spezifischer

Lernmaterialien aus existierenden Fragmenten (z.B. Texten, Bildern oder Videos). Das *Connexions*-Projekt dient der Erstellung von Lerninhalten sowie der Entwicklung einer Infrastruktur hierfür (z.B. durch entsprechende Tools). Dazu wird die eigens entwickelte Beschreibungssprache CNXML benutzt [27]. Im Gegensatz zu LEBONED gehen diese beiden Projekte jedoch von einer vollständigen Neuerstellung der Lernmaterialien aus.

Auf dem Gebiet der (semi-)automatischen Dokumentenanalyse gibt es zahlreiche Ansätze, die im Gegensatz zu den bei uns angewandten Algorithmen jedoch oftmals auf der Analyse von Bildern eines Dokuments bestehen (z.B. einem Bitmap der Seite eines PDF-Dokuments). Ein Überblick über Algorithmen zur Erkennung von Dokumentenlayouts aus solchen Bitmaps gibt [28], während die Erkennung logischer Strukturen von Dokumenten aus Bitmaps in [29] beschrieben wird. Als Ansatz für die Dokumentenanalyse basierend auf einer Seitenbeschreibungssprache sei AIDAS genannt, das sich mit der Erkennung logischer Strukturen in PDF-Dokumenten befasst [30].

5 Zusammenfassung und Ausblick

In diesem Papier wurden Forschungsergebnisse des Projektes LEBONED präsentiert, dessen Ziel die Entwicklung einer möglichst allgemeingültigen Lösung für die Integration digitaler Bibliotheken in webbasierte LMS darstellt. Auf recht abstrakter Ebene wurde dazu zunächst beschrieben, welche Modifikationen an einer Standardarchitektur für herkömmliche LMS vorgenommen werden müssen. Das anschließend beschriebene Modell für den Import und die Weiterverarbeitung von Dokumenten und zugehörigen Metadaten stellt einen bedeutsamen Ausschnitt dieser Architektur dar. Insbesondere die Entwicklung des Metadatenformats lxSCORM ist hierbei ein wichtiges Forschungsergebnis, da dieses die besonderen Anforderungen bei der Beschreibung und Verwendung von Dokumenten aus digitalen Bibliotheken in LMS berücksichtigt.

Die vorgestellten Softwarewerkzeuge Learning Object Creator und Learning Object Composer zur Fragmentierung monolithischer Dokumente in kleinere Lernobjekte und zur Komposition solcher Lernobjekte ermöglichen die effiziente und flexible Erstellung neuer Lernmaterialien auf Basis existierender Dokumente. Sie bilden deshalb wesentliche Bausteine für die angemessene (Wieder-) Verwendung importierter Materialien und runden die hier präsentierte Gesamtinfrastruktur somit sinnvoll ab.

Literatur

1. Schulmeister, R., Virtuelle Universität Virtuelles Lernen. 2001: Oldenbourg Verlag.
2. Rosenberg, M.J., e-Learning - Strategies for Delivering Knowledge in the Digital Age. 2001: McGraw-Hill.
3. Encarnação, J., W. Leidhold, and A. Reuter, Szenario: Die Universität im Jahre 2005, in Studium online: Hochschulentwicklung durch neue Medien, B. Stiftung and H.N. Stiftung, Editors. 1999.
4. Dodds, P., Sharable Content Object Reference Model (SCORM) - Version 1.2 - The SCORM Overview. 2001, Advanced Distributed Learning (ADL).
5. Baumgartner, P., H. Häferle, and K. Maier-Häferle, E-Learning Praxishandbuch - Auswahl von Lernplattformen. 2002: StudienVerlag.
6. Arms, W.Y., Digital Libraries. 2000: MIT Press.
7. Endres, A. and D.W. Fellner, Digitale Bibliotheken - Informatik-Lösungen für globale Wissensmärkte. 2000: dpunkt.verlag.
8. Lesk, M., Practical Digital Libraries - Books, Bytes & Bucks. 1997: Morgan Kaufmann Publishers.
9. Oldenettel, F. Das eVerlage-System: Eine digitale Bibliothek für kostenpflichtige wissenschaftliche Verlagsprodukte. in Leipziger Informatik Tage (LIT) 2002. 2002. Leipzig: infix Verlag.
10. Ober, J., The California Digital Library. D-Lib Magazine, 1999. 5(3).
11. Witten, I.H., S.J. Cunningham, and M.D. Apperley, The New Zealand Digital Library Project. D-Lib Magazine, 1996. 2(10).
12. Bainbridge, D., D. McKay, and I.H. Witten, Greenstone Digital Library - Developer's Guide. 2001, University of Waikato, New Zealand.
13. Universitätsbibliotheken, A.d. and A.d. Fachhochschulbibliotheken, Die Digitale Bibliothek NRW - Konzept. 1998.
14. Knop, J. and W. Haverkamp, E-Business in der Hochschule: Wirklichkeit, Vision und Voraussetzungen, in Strategien und Prozesse für neue Geschäftsmodelle - Praxisleitfaden für E- und Mobile Business, J. Frischmuth, W. Karrlein, and J. Knop, Editors. 2001, Springer Verlag. p. 81 - 98.
15. Initiative, D.C.M., Dublin Core Metadata Element Set, Version 1.1: Reference Description. 1999, Dublin Core Metadata Initiative.
16. Archivists, S.o.A., Application Guidelines for Version 1.0 - Encoded Archival Description (EAD) - Document Type Definition (DTD), Version 1.0 - Technical Document No. 3. 1999, Society of American Archivists.
17. Furrie, B., Understanding Marc Bibliographic: Machine-Readable Cataloging. 6th edition (January 2001) ed. 2001: Library of Congress.
18. Learning Technology Standardization Committee, I., Draft Standard for Learning Object Metadata. 2001, Institute of Electrical and Electronics Engineers, Inc.: New York.

19. Hammer, J., et al. Template-Based Wrappers in the TSIMMIS System. in ACM SIGMOD International Conference on Management of Data. 1997. Tucson, Arizona, USA: ACM Press.
20. Pulkowski, S. Intelligent Wrapping of Information Sources Getting Ready for the Electronic Market. in 10th VALA Conference on Technologies for the Hybrid Library. 2000. Melbourne, Australia.
21. Dodds, P., Sharable Content Object Reference Model (SCORM) - Version 1.2 - The SCORM Content Aggregation Model. 2001, Advanced Distributed Learning (ADL).
22. Oldenettel, F. and M. Malachinski. The LEBONED Metadata Architecture. in 12th International World Wide Web Conference. 2003. Budapest, Hungary.
23. Gövert, N., N. Fuhr, and C.-P. Klas. Daffodil: Distributed Agents for User-Friendly Access of Digital Libraries. in European Conference on Digital Libraries (ECDL). 2000.
24. Fischer, S., Course and exercise sequencing using metadata in adaptive hypermedia learning systems. *Journal of Educational Resources in Computing (JERIC)*, 2001. 1(1es): p. 5.
25. Pöyry, P., K. Pelto-Aho, and J. Puustjärvi. The role of metadata in the CUBER system. in Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology. 2002. Port Elizabeth, South Africa: South African Institute for Computer Scientists and Information Technologists.
26. Consortium, I.G.L., IMS Digital Repositories Interoperability - Core Functions Information Model. 2002, IMS Global Learning Consortium, Inc.
27. Radaelli-Sanchez, R., The Basic CNXML. 2002, The Connexions Project.
28. Okun, O., D. Doermann, and M. Pietikäinen. Page segmentation and zone classification: a brief analysis of algorithms. in International Workshop on Document Image Analysis and Understanding. 2001. American University in Dubai.
29. Summers, K., Automatic Discovery of Logical Document Structure. 1998, Cornell Computer Science Department.
30. Anjewierden, A. AIDAS: Incremental logical structure discovery in PDF documents. in 6th International Conference on Document Analysis and Recognition (ICDAR). 2001. Seattle.

Kontakt

Frank Oldenettel, Michael Malachinski
Oldenburger Forschungs- und Entwicklungsinstitut für
Informatik-Werkzeuge und –Systeme (OFFIS)
Escherweg 2, 26121 Oldenburg
{Frank.Oldenettel, Michael.Malachinski}@offis.de

MPRESS - transition of metadata formats¹

Judith Plümer

Universität Osnabrück, Fachbereich Mathematik/Informatik

Abstract

Progress in science is based on the availability and accessibility of its state of the art. Nowadays, in mathematics and related sciences this state of the art is reflected in electronically published preprints. In this way every author becomes his or her own publisher, consequently there is a lack of an organized accessibility of this material.

The Mathematics PREprint Server System (MPRESS) is a service which offers an index of mathematical preprints which have been electronically published all over the world.

Referencing more than 50,000 documents which reside on more than 100 WWW servers, MPRESS is the largest index of mathematical preprints worldwide. The index data of the preprints are enriched by the use of metadata which improve the quality of retrieval dramatically.

The development of metadata schemes and metadata encoding enforces a transition from plain metadata to RDF, as recommended by W3C.

The paper sketches out the system of MPRESS and the related development in the area of metadata. The problems that these developments cause for a system like MPRESS are discussed and the transition to a new metadata scheme is presented together with the improvements that result from this transition.

MPRESS - Transformation von Metadaten Formaten

Wissenschaftlicher Fortschritt basiert auf der Verfügbarkeit des State of the Art. In Mathematik und verwandten Wissenschaften spiegelt sich der State of the Art in Preprints wieder, die heutzutage elektronisch publiziert werden. Da jeder Autor so zum Verleger wird, fehlt es an Organisation zur Verfügbarkeit des Materials.

Das Mathematical Preprint Server System (MPRESS) ist ein System zur Bereitstellung eines Index mathematischer Preprints aus der ganzen Welt.

Mit über 50.000 Dokumenten von über 100 WWW-Servern ist MPRESS der größte Nachweis mathematische Preprints weltweit. Die Indexdaten der Pre-

¹ This project was funded by the Deutsche Forschungsgemeinschaft (DFG).

prints werden durch die Nutzung von Metadaten verbessert. Die qualitative Suche wird auf diese Weise sehr effektiv.

Die Entwicklung von Metadaten Schemata und Metadaten Codierung zwingt uns nun vom flachen HTML-META tag zum detaillierten Resource Description Framework (RDF) überzugehen, wie es vom W3C empfohlen wird.

Der Artikel skizziert MPRESS und die Entwicklung der Metadaten Codierung. Die Probleme, die sich für ein System wie MPRESS aus einer Umstellung des zugrundeliegenden (Meta)Datenformates ergeben, werden diskutiert und Lösungsansätze sowie die resultierenden Verbesserungen werden vorgestellt.

Introduction

For a long time mathematicians used to circulate preprints and exchange ideas within small circles.

Today, however, mathematicians and other scientists put their scientific results on departmental WWW-Servers in forms of preprints, or they send it to preprint servers like arXiv.org to avoid the long period of time between the submission of an article and its publication in a scientific journal.

However, it soon became obvious that this large amount of data was hardly useful just as a set because people only knew how to find their own preprints. But how could he or she find preprints of a colleague at university X or preprints on a certain topic?

Large search engines in the internet, like Alta Vista or Google, are not well adapted to finding specific scientific material because scientific material only covers a very small part in the set of all internet documents and resources.

What today's mathematicians need in order to retrieve scientific information from the web is a *specific way of searching* for information. This is provided by metadata and their *specific ways of navigating*. Part of this is given by the Mathematics Subject Classification (MSC).

The Math-Net project in Germany [Dalitz, 1997] was designed to answer these needs and it was conceived as a first step in order to design and implement a worldwide electronic information and communication system for mathematics and to turn Math-Net into an international endeavor.

MPRESS, which is supported by the Math-Net Project, started under the auspices of the European Mathematical Society. The CEIC (Committee on Electronic Information and Communication), which is established by the International Mathematical Union (IMU), has designated MPRESS as a Math-Net service.

In 1995 the Dublin Core Initiative proposed the formal description of documents by using the Dublin core 15 elements. The idea was to store these metadata in the head of HTML files by using the HTML META tag. MPRESS was

the first distributed service that made use of this idea in order to improve retrieval.

Nowadays the W3C recommends the use of RDF to express descriptions of documents. This enforces systems like MPRESS to change their architectures in order to be able to handle RDF and former formats.

The paper describes the architecture of MPRESS and the philosophy of RDF in comparison with the HTML META tag, in this way it explains the difficulties that arise during the transition to RDF and how these difficulties are being handled.

Harvest working with plain metadata

The system MPRESS/Math-Net.preprints is an index for mathematical preprints, it is not a full text archive. The idea is to create an index of mathematical preprints without storing them on one central preprint server.

Usually central archives offer search mechanisms, but the preprints on departmental servers are at most searchable via Web-Robots of the Alta Vista type. Since the number of scientific mathematical material is very small it is nearly impossible to find mathematically relevant papers in the Web via those general search engines. A solution to this problem is to generate an index of the distributed documents. MPRESS/Math-Net.preprints is such a system. Currently MPRESS compiles links to about 50,000 mathematical preprints in the world. The quality of the retrieval in MPRESS is mainly founded on the good use of metadata.

HTML Metadata in harvest

The metadata set is based on the metadata set that the American Mathematical Society (AMS) created for their preprint server (<http://www.ams.org/preprints>). The AMS stored these metadata in ASCII format without using any particular scheme.

In MPRESS we store these metadata inside HTML documents that serve as title pages for the preprints. Obviously it is necessary to offer information like "author" and "title" in a way that is independent from the subject "mathematics". The Dublin Core initiative [Kunze, 1999] offers such a scheme which is user-friendly and allows subject specific extensions.

A full list of the used items and the current encoding of these items can be found at <http://www.math-net.de/project/metadata/preprint.html>, as there are the name of the author, the email address of a contact person, Mathematical Subject Classification (MSC) codes, an abstract, notes, the date of creation and last

modification, the status of the document (preprint/article), the mimetype of the original file, its language, source, URL and a rights statement.

The first extension to the standard Dublin Core element set that is specific within the set was the use of „DC.subject.msc“ to store the MSC codes.

The encoding of the metadata is currently done via the HTML2.0 META tag. However, the encoding of metadata in HTML 2.0 leads to some problems. For example there is no grouping mechanism which says for example that this email address belongs to that person's name. That is the reason why we store only one email address in the metadata files. HTML4.0 gives more attributes to the META tag, but the grouping problem is not solved there.

Another problem with the META tag in HTML is the double storage of data, once to be found in the head as metadata and again in the body of the document as a visible text for human eyes. If someone changes the visible part of the metadata document it seems to be correct for him in the browser, but it may contain incorrect metadata. These problems can be solved when switching to RDF by using DC and vCard vocabulary and object typing as we will describe below.

As already mentioned, the metadata files or title pages of the preprints contain an abstract and MSC codes. Therefore the creation of the files should be done by the authors of the papers themselves. To do so a detailed knowledge about Dublin Core and HTML encoding is necessary or a tool that cares for the syntactical correctness. Such a tool, called the Mathematics Metadata Markup Editor, was developed at Osnabrück in collaboration with E. Hilf, Th. Severiens, M. Jost and M. Kaplan.

This tool has been successively enhanced to introduce features such as controls on the input, MSC browsing, and the output of Dublin Core metadata encoded in HTML in a first version. The current version additionally supports RDF.

The author of a paper can type in the respective metadata and via mouse click the metadata file is generated by a perl script. The Mathematics Metadata Markup Version 3.1 tool can be used remotely or downloaded for local installation from <ftp://ftp.math.uos.de/pub/MMM/>.

Now the stored metadata guarantee a much higher quality in retrieval than for example a full text search on the original documents. On the one hand the result sets are more specific and on the other hand the results can be presented in a well structured way.

How documents enter the system and how they are presented

Once the metadata are stored on a WWW server they must be collected in some way. However, the whole document cannot be stored in a central index. The copyright of the authors allows the storage of copies for private or personal use, but it is forbidden to keep them in a database and distribute them without the permission of the author.

The gathering part of the Harvest [Hardy, 1996, Technical Report] software which is used in Math-Net is responsible for this collecting process. It is a configurable robot. Now the gathering of the documents does not mean making copies and keeping them somewhere. It means taking a temporary copy, extracting the relevant information from this copy, storing this information in a database and deleting the temporary copy.

One gathering agent is determined by several configuration files that have to be created by an administrator. There is one configuration file that contains the URLs that the agent should gather and evaluate together with a lot of parameters which say whether he should run recursively, at which depth the recursion should stop, how many documents should be gathered at most, which documents should be included or excluded during the recursion, how many different hosts may be visited during the recursion and which types of protocols should be used (ftp, http).

The other configuration files can be used in their default form or be modified, too. As there are configuration files which say how to handle different mimetypes and formats, how to interpret suffixes, and how to summarize different formats.

The gatherer pipes the retrieved document through the essence machine [Hardy, 1996, Trans. Comp. Sci] that generates summaries of the documents in SOIF [Hardie, 1999] (summary object interchange format, see below). The SOIF documents are stored by the gatherer and the original resources are deleted. The SOIF documents of one agent reside in a gnu-zipped ASCII file with some additional information. For example, there is a small databases storing the MD5 [Rivest, 1992] numbers of the original files to avoid multiple copies of one document. Internally the MD5 number is also used to check whether a document has been changed or not in case the agent does not visit a site for the first time. (The agents of MPRESS run every other week.) If the document has not been changed, the essence process does not have to be run again on the same document. In this case only the time-to-life (TTL) of the respective SOIF-record is modified. This saves local computing power, but it does not save netload, which would be desirable. So what we wanted was an incrementally running agent. The original harvest software was modified in this way so that it runs incrementally for MPRESS.

There are summarizers for a variety of formats. The pages in the Web that contain mathematically relevant information are usually stored in PostScript, PDF or HTML. The summarizers of the harvest software do not handle 8-bit and unicode characters correctly. So we had to modify the PostScript and HTML summarizer to interpret at least Umlauts and „ß“ for German needs in order to allow correct responses when querying for German names. This problem is solved with X-Harvest and HyREX because they are able to handle Unicode characters (see below).

The summarizers generate SOIF records for each document. SOIF was designed for the storage and exchange of the summaries of documents which may originate in different formats. It was built as a part of the harvest software. The structure of SOIF is simple and effective. A SOIF object consists of multiple attribute-value pairs. For example

author{ 16}: Erwin Mustermann

is such a pair, the number in brackets gives the number of characters of the value.

There are no restrictions on the number and names of attributes.

This principle of attribute-value pairs makes SOIF equivalent to the use of the HTML2.0 META tag with its NAME-CONTENT pairs. But it can also be used to store information from the body of a HTML document or from other formats.

For example, the HTML summarizer of harvest assigns all words between `<Hn>` and `</Hn>` to the header attribute.

Obviously, it makes sense to build such attribute-value pairs from any markup language like HTML, SGML, XML or TeX/LaTeX.

The summarizer for documents in a typeset format as .dvi, .ps or .pdf work differently, they only store the first 100 words of these documents as keywords because the text is not semantically structured. One could think of criteria to extract more structure out of these unstructured documents via methods of artificial intelligence, heuristics or the use of thesauri (see below).

The summarizers of harvest also store metadata such as the mimetype of the original document, its URL, URLs of links contained in the document, name and host of the gathering agent, time of generation, update time and time to live. All these pieces of information play an important role for the system because the information is not constant. They are used to control the expiration times.

Collection of documents and problems of heterogeneity

Since SOIF objects are just ASCII files, they can be transferred by any protocol as SOIF streams. Harvest builds a database or an index of SOIF objects by using Glimpse which makes the attribute-value pairs searchable.

Now the problem arises when upgrading to HTML4.0 or to XML with metadata in RDF (Resource Description Framework). In MPRESS we currently store SCHEME information of HTML4.0 META as a qualifier. For example

`<META NAME="DC.subject" SCHEME="msc91" VALUE="19D10">`

is mapped to

DC.subject.msc{5}: 19D10

This is done by a modification of harvest's HTML summarizer. But now under Glimpse you won't get this object when searching for `DC.subject=19D10`.

This does not create any problems in a controlled environment, but when you start to import SOIF documents from remote sites your metadata become heterogeneous.

This problem currently occurs in MPRESS with the metadata files we get from the arXiv.org mirror at Augsburg: The name of the author of a paper is stored there as `DC.creator` and not as `DC.creator.PersonalName` as inside of MPRESS. This problem of heterogeneity is solved during the import of the SOIF documents into the broker. The gatherer command that manages this import pipes the SOIF objects through a Perl script that substitutes the older DC coding of the author's name with the newer version.

In analogue to the SCHEME problem the handling of sub qualifiers in DC coded in RDF is difficult at this stage.

After collecting the documents and generating SOIF objects for them, each gathering agent exports the SOIF objects via a daemon. Now the broker component collects the SOIF objects from several daemons and computes an index of these preprints, or more exactly an index of their SOIF objects. The gatherers that feed MPRESS are also distributed over the world which makes an upgrading of software or used formats difficult. The user of MPRESS can only see the broker which is an interface for searching in this information pool.

Not all stored metadata elements can be used and evaluated by the user. Currently searches for authors, titles and keywords (and dates in France) are supported, but if you know the correct syntax of the other metadata items, you can also use the full text search field to query for the other items.

An important feature of MPRESS is the browse function besides the search functionality. Here the user gets a table for the several levels of the MSC and the number of preprints classified in the respective categories. Then he can browse through the next level of the MSC or search in the respective subset of preprints.

To achieve such a functionality of the broker, one has to refine the perl scripts of the harvest distribution which organize the preparation of the query to the glimpse database and which arrange the answers from the glimpse database to be nicely displayed. In detail this is the `BrokerQuery.pl` script in `harvest1.4` or the `nph-search` script in `harvest1.5`.

Especially the browse functionality is not part of the harvest software. The pages that appear during the browsing process are generated pseudo-automatically: The pages reside in the file system of the WWW server, but they are generated once a day to be up to date.

This is done by a cronjob, which is a perl script that queries the broker for the different MSC categories. Analogously the MSC specific search interfaces are generated on the fly to send an MSC enriched query to the broker that filters out the papers without the correct MSC codes. That means the browsing function is

based on a few perl scripts that have no influence on the performance of the machine. If someone ran an extra broker for each MSC classification, the performance of the system would decrease drastically.

Obviously the broker component is independent of the used metadata scheme and the offered kind of material which are preprints in this example. The same methods of retrieval work for any kind of documents enriched with metadata. .

In the following we will discuss the advantages of RDF as a format for storing metadata in MPRESS and how the system can migrate.

RDF compared to HTML META

We don't want to discuss the general advantages of the Resource Description Framework RDF (<http://www.w3.org/RDF/>) [Brickley, 2001] here. We just want to give an impression of what the difference implies for a system like MPRESS.

RDF stores information in units of statements consisting of Subject, Predicate and Object. The combination of statements results in complex graphs that can be expressed in RDF. HTML META stores only attribute/value or NAME/CONTENT pairs. Or in the notion of RDF it stores Predicates and Objects so that the subject is thought of being the document itself.

An example of the difference between HTML META and RDF in MPRESS is the presentation of two authors of a paper and their email addresses.

In HTML this would look like:

```
<meta name="DC.Creator.PersonalName" content="Mustermann, Marie"/>
<meta name="DC.Creator.Email" content="marie@here.org"/>
<meta name="DC.Creator.PersonalName" content="Musterfrau, Paul"/>
<meta name="DC.Creator.Email" content="paul@there.org"/>
```

In the XML presentation of RDF it looks like:

```
<dc:creator>
  <rdf:Bag>
    <rdf:li>
      <mn:Person vCard:EMAIL="marie@here.org">
        <rdfs:label rdf:parseType="Literal">Marie
                                Mustermann</rdfs:label>
      </mn:Person>
    </rdf:li>
    <rdf:li>
      <mn:Person vCard:EMAIL="paul@there.org">
```

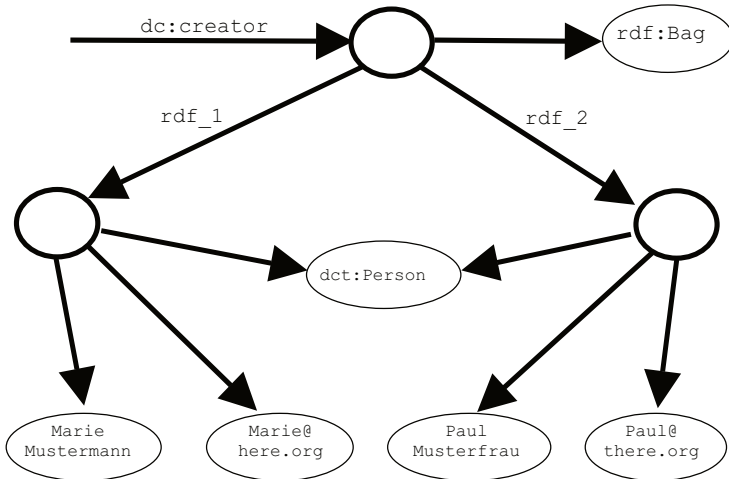
```

<rdfs:label rdfs:parseType="Literal">Paul
                                Musterfrau</rdfs:label>

</mn:Person>
</rdf:li>
</rdf:Bag>
</dc:creator>

```

The graph presentation of RDF is easier to read for human eyes:



The problem with HTML META is the missing function of bracing elements together and of giving nested statements.

A full description of the RDF metadata that is used in MPRESS and in the respective MMM tool was worked out by the MathNet technical advisory board of the International Mathematical Union and is available at <http://www.iwi-iuk.org/material/RDF/1.1/profile/MNPreprint/>.

In the next section we want to discuss the problems that arise when one wants to migrate from the current system of MPRESS with use of HTML META to a system based on RDF which we are currently doing in a project funded by the Deutsche Forschungsgemeinschaft (DFG).

Migration from HTML META to RDF

Above we indicated the advantages that RDF has in comparison with HTML META. The question that we want to discuss here is how the software can handle the richer structure.

HTML META is in a mathematical sense equivalent to the SOIF format since both store attribute/value pairs. But that precludes the use of RDF on the base of the current harvest software. What we need is a software with the power to handle RDF and the wisdom to integrate SOIF data that come from some gatherer agents in the world and don't want to update their software.

In the CARMEN (<http://www.math.uos.de/projects/carmen/>) project of the federal ministry of science (Global Info program of the BMBF) tools were developed that can be plugged together to operate MPRESS on the basis of RDF.

X-Harvest

The X-Harvest [Kokkelink, 2000] software is a modification of the Harvest-NG (<http://webharvest.sourceforge.net/ng/>) software that substitutes the internal SOIF format by RDF. That means X-Harvest is a substitute for the gatherer component of the Harvest software which stores the summaries of the documents in RDF format. X-Harvest is completely written in Perl. That makes addition and modification of features easy but on the other hand it makes the installation a real challenge: the Harvest-NG software needs the installation of 8 additional Perl modules which require further Perl modules. X-Harvest needs additional Perl modules which piles up to 25 Perl modules totally that have to be installed. Therefore we decided to implement a script that takes care of the installation procedure.

But this modular architecture solves for example the problem of character sets because X-Harvest uses the Unicode module and hence handles Unicode characters.

As well there are modules for the X-Harvest software that are able to solve the heterogeneity problems that we mentioned above:

What do we do with HTML documents without any metadata and with documents coming up in formats that are not able to carry metadata information like PostScript? For this purpose there is a summarizer that generates metadata out of HTML documents on probabilistic guesses and a summarizer for PostScript documents that extracts metadata by given heuristics (<http://www.math.uos.de/projects/carmen/API11/>). The use of heuristics in this context makes complete sense since mathematical papers are at most all of the same structure and mostly offered in PostScript which was generated out of TeX/LaTeX. This, however, is an aspect for which the described solution does not have to be scalable for other disciplines.

Transition to RDF creates the heterogeneity problem that we now have, which is a zoo of sources offering index data for MPRESS: old harvest gatherers offering SOIF, X-Harvest gatherers offering RDF and agents that export index data in the Open Archives Protocoll (<http://www.openarchives.org/>). But this is unavoidable as long as one accepts data coming up in the former formats what we want to do in MPRESS.

We want to import this zoo of sources into HyREX [Gövert, 2002]. During the CARMEN project a script (`broker.pl`) was developed that collects data from harvest gatherers, X-Harvest gatherers and interfaces using the Open Archives Protocoll. The data that are stored by this script have to be converted to XML using for example `soif2xml`. Then the whole set of data is in XML format and converted again using XSLT. This procedure results in a set of XML documents that are valid against the DTD we are using to index the data with HyREX.

Bibliography

- Brickley, Dan; Bray, Tim: "What is RDF?",
<http://www.xml.com/pub/a/2001/01/24/rdf.html>, 2001.
- Dalitz, Wolfgang; Grötschel, Martin; Lügger, Joachim: "Information Services for Mathematics in the Internet (Math-Net)", 15th IMACS World Congress 1997 on Scientific Computation, Medelling and Applied Mathematis, Volume 4: Artificial Intelligence and Computer Science, Wissenschaft und Technik, Berlin, 1997, p. 773-778.
- Gövert, Norvert; Großjohann, Kai: „HyREX Manual“, <http://www.is.informatik.uni-duisburg.de/projects/hyrex/manual.pdf>.
- Grötschel, Martin; Lügger, Joachim: "Scientific Information Systems and Metadata", Classification in the Information Age, Springer, 1999, p.3-20.
- Hardie, T.; Bowman, M.; Hardy, Darren R.; Schwartz, Michael F.; Wessels, Duane: "CIP Index Object Format for SOIF Objects", IETF RFC2655, 1999.
- Hardy, Darren R.; Schwartz, Michael F.: "Customized information extraction as a basis for resource discovery", ACM Transactions on Computer Systems, Vol. 14 no. 2, 1996, p 171-199.
- Hardy, Darren R.; Schwartz, Michael F.; Wessels, Duane: "Harvest – Effective use of Internet Information", University of Colorado at Boulder, Technical Report CU-CS-743-94, 1996.
- Kokkelink, Stefan: "Simple XML/RDF extension of Harvest-NG",
<http://www.math.uos.de/projects/carmen/AP7/DOM/dom.htm>, 2000.
- Kunze, John: "Encoding Dublin Core in HTML", IETF RFC 2731, 1999.
- Rivest, R. L.: "The MD5 Message-Digest Algorithm", IETF RFC 1321, 1992.

Contact

Judith Plümer
Universität Osnabrück
Fachbereich Mathematik/Informatik
49069 Osnabrück
Germany
Tel: +49 541 969 2526
Fax: +49 541 969 2770

Einsatz von Lernmanagementsystemen in traditionellen Lehrveranstaltungen am Beispiel der Universitäten Oldenburg und Osnabrück

Ein Erfahrungsbericht

Dennis Reil

OFFIS

Abstract

This paper gives an overview about the implementation of an e-learning environment of the ELAN-pilot Oldenburg and Osnabrück (epolos). It shows the selection of a learning platform and the use in traditional courses of the participating universities. It further outlines the support of teachers especially by the development of reference models for traditional courses.

1 Einleitung

Das *e-learning academic network Niedersachsen* (ELAN¹) des niedersächsischen Ministeriums für Wissenschaft und Kultur (NMWK) hat den nachhaltigen Aufbau einer e-Learning-Infrastruktur an niedersächsischen Universitäten zum Ziel. Es wurden drei Pilotprojekte gebildet, die unter anderem den Einsatz von so genannten Lernmanagementsystemen erproben. Eines dieser Pilotprojekte ist der *ELAN-Pilot Oldenburg Osnabrück* (epolos²).

Das *Content-Engineering Labor* (CELab³) hat als Leitprojekt von epolos zwei Lernmanagementsysteme beschafft und setzt sie für die Lehre an den Universitäten Oldenburg und Osnabrück ein.

Dieser Beitrag gibt ausgehend vom Begriff des Lernmanagementsystems (LMS) einen kurzen Überblick über den zeitlichen Verlauf der Einführung an den Universitäten Oldenburg und Osnabrück. Anschließend werden der Auswahlprozess in epolos sowie die aufgestellten Kriterien vorgestellt. Des Weiteren werden Maßnahmen zur Einführung des Lernmanagementsystems und Unterstützungsleistungen für Lehrende erläutert. Hierzu gehören Referenzmodelle

1 <http://www.elan-niedersachsen.de/>

2 <http://www.epolos.de/>

3 <http://www.celab.de/>

für traditionelle Lehrveranstaltungen. Im Folgenden wird zunächst ein kurzer Überblick über das Projekt ELAN und seine Ziele gegeben.

2 Übersicht ELAN / epolos

Das e-Learning academic network Niedersachsen (ELAN) verfolgt als Hauptziel den Aufbau einer nachhaltigen e-Learning-Infrastruktur an niedersächsischen Universitäten. Hierzu wurden drei Pilotprojekte gebildet:

- Hannover / Braunschweig⁴
- Göttingen / Clausthal⁵
- Oldenburg / Osnabrück (epolos)

„Im ELAN-Pilot Oldenburg/Osnabrück - kurz epolos - arbeiten die Universitäten Oldenburg und Osnabrück gemeinsam am Aufbau eines Netzwerkes mit dem Ziel, den Einsatz von e-Learning-Angeboten für die Aus- und Weiterbildung an Hochschulen voranzutreiben.“ (epolos-Antrag).

Der Leitgedanke von *epolos* ist das Content-Engineering. Als Leitprojekt ist das Content-Engineering Labor (CELab) für Beschaffung, Betrieb und Erprobung von Lernmanagementsystemen zuständig. Zum derzeitigen Zeitpunkt (Mai 2003) betreibt das Content-Engineering Labor zwei Lernmanagementsysteme:

- Blackboard Learning System 5.7 ML und
- ILIAS

Bei Blackboard handelt es sich um ein kommerzielles Lernmanagementsystem, während ILIAS ein Vertreter der Open Source-Systeme ist. Daneben wird an der Universität Osnabrück ein weiteres kommerzielles System (WebCT) betrieben. Diese drei Systeme stehen Lehrenden an beiden Universitäten zum Einsatz in der Universitätslehre bereit. Wie stark und in welchem Maße dies auch tatsächlich genutzt wird, zeigt der letzte Abschnitt dieses Beitrags.

Zunächst soll eine kurze Historie des CELabs den Ablauf der Einführung eines Lernmanagementsystems an der Universität Oldenburg dokumentieren. Das CELab wurde im Sommer 2002 an der Universität Oldenburg gegründet. Von August 2002 an wurde die Auswahl eines kommerziellen Lernmanagementsystems durchgeführt und im September ein solches beschafft. Daraufhin wurde Anfang Oktober die Installation und Inbetriebnahme des Systems vorgenommen und dieses auf einem Multimediatag der Universität vorgestellt. Im Wintersemester 2002/2003 erfolgte dann ein erster praktischer Einsatz an den beiden Universitäten Oldenburg und Osnabrück.

4 <http://www.learninglab.de/elan/index.html>

5 <http://www.elan.tu-clausthal.de/>

Parallel dazu wurde durch das VirtuOS⁶ (Zentrum zur Unterstützung virtueller Lehre der Universität Osnabrück) das Lernmanagementsystem WebCT beschafft und installiert. Dieses wurde aber vorwiegend nur in der Osnabrücker Lehre getestet, da WebCT für die Universität Oldenburg aufgrund der zu komplexen Bedienung bereits in der Vorauswahl ausgeschieden war.

3 Auswahl von Lernmanagementsystemen

Der Begriff des Lernmanagementsystems wird häufig unterschiedlich definiert. Daher soll im Folgenden zunächst eine Definition erfolgen.

3.1 Begriff des Lernmanagementsystems

Laut Baumgartner (2002) ist der Begriff des Lernmanagementsystems synonym zum Begriff des Lernportals und der Lernplattform. Eine Lernplattform wird wie folgt definiert:

„Unter einer webbasierten Lernplattform ist eine serverseitig installierte Software zu verstehen, die beliebige Lerninhalte über das Internet zu vermitteln hilft und die Organisation der dabei notwendigen Lernprozesse unterstützt“ ([1], S.24).

Diese Definition sowie die sehr ausführliche Evaluation ([1]) sind Grundlage des Auswahlprozesses.

3.2 Auswahlprozess

Zu Beginn eines Auswahlprozesses steht die Gewinnung einer Marktübersicht. Auf dem Markt der Lernmanagementsysteme gibt es derzeit einige hundert verschiedene Systeme zur Auswahl. Davon sind einige als Open Source-Systeme kostenlos verfügbar. Zu dieser Klasse gehören z.B.:

- Claroline⁷
- ILIAS⁸
- Moodle⁹
- Stud.IP¹⁰

Es ist festzustellen, dass auf dem Markt mehr kommerzielle Lernmanagementsysteme als solche aus dem Open Source-Bereich vorhanden sind. Laut Baum-

6 <http://www.virtuos.uni-osnabrueck.de/>

7 <http://www.claroline.net/>

8 <http://www.ilias.uni-koeln.de/ios/>

9 <http://www.moodle.org/>

10 <http://www.studip.de/>

gartner (2002) ist daraus abzulesen, dass es eine sehr komplexe Aufgabe ist, ein Lernmanagementsystem zu entwickeln. Insbesondere die Anforderungen hinsichtlich Performance und Verfügbarkeit führten dazu, dass es mehr kommerzielle Anwendungen gebe.

Neben der Marktübersicht ist eine gute Übersicht über verfügbare Evaluationen bzw. Erfahrungsberichte aus dem Bereich der Lernmanagementsysteme essentiell. Hier sind insbesondere die Publikationen von Baumgartner (2002) und Schulmeister (2000) zu nennen. Daneben gibt es einige Publikationen zu grundsätzlichen Anforderungen an eine e-Learning Plattform ([2]; [5]).

Zu Beginn des Auswahlprozesses steht die Entscheidung, ob auf bestehende Studien zurückgegriffen werden sollte oder ob eine eigene Evaluation der am Markt verfügbaren Systeme durchgeführt werden sollte. Für eine eigene Untersuchung spricht, dass dann nach eigenen Kriterien untersucht werden kann. Daneben gibt es allerdings eine ganze Reihe von Gründen, keine eigene Evaluation durchzuführen. Insbesondere der hohe Zeitaufwand, der hierfür nötig ist, lässt oft keine eigene Untersuchung zu. Hinzu kommt das Problem, Evaluationsversionen der Systeme zu erhalten. Die Hersteller halten sich hier oft sehr bedeckt und stellen nur Demosysteme zum Test bereit. Begründet wird dies mit dem häufig hohen Aufwand für die Installation eines Systems und mit dem für den Betrieb und die Administration nötigen, jedoch fehlenden Know-how. Dies ist ein berechtigter Grund, denn ein Lernmanagementsystem ist, wie bereits ausgeführt, ein sehr komplexes System, das aus diversen zusammenarbeitenden Komponenten besteht. Daher ist es häufig nicht einfach zu installieren. Die bereitgestellten Demosysteme haben zudem meistens den Nachteil, dass man keine Administratorenrechte besitzt und somit Einstellungen nur auf Kursebene ändern kann. Somit lassen sich die Möglichkeiten hinsichtlich Adaption, Administration, etc. nicht ausreichend erproben. Weiterhin fehlt auch der Einblick in die Probleme, die bei einer Installation auftreten können und darüber, welche Hardware ein solches System effektiv benötigt. Dies wird für die Demosysteme nicht selten verschwiegen. Als letztes und vermutlich wichtigstes Argument gegen eine eigene Evaluation steht die starke Fluktuation der Hersteller bzw. Systeme am Markt. Ist eine solche zeitaufwendige Untersuchung endlich durchgeführt, so gibt es möglicherweise den Anbieter und das System gar nicht mehr, das am besten bei der Evaluation abgeschnitten hat. Des Weiteren kann es derweil wieder neue Versionen geben, die Nachteile alter Versionen ausgleichen oder aber sogar zu neuen Nachteilen führen. Aus diesem Grunde wurde für das Projekt epolos ein anderes Vorgehen gewählt, das im folgenden Abschnitt beschrieben wird.

3.3 Der Auswahlprozess im Projekt epolos

Grundlage des Auswahlprozesses von epolos waren bestehende Evaluationen, insbesondere die von Baumgartner (2002) und Schulmeister (2000), die der Vorauswahl bzw. der Einschränkung der genauer zu untersuchenden Lernmanagementsysteme dienten. Folgende Systeme wurden auf dieser Grundlage für eine nähere Untersuchung ausgewählt:

- Blackboard¹¹
- CliX¹²
- Hyperwave¹³
- Lotus Learning Space
- Saba
- WebCT¹⁴

Vor der Untersuchung dieser Systeme wurden Anforderungen definiert, die für das epolos-Projekt als wichtig erachtet wurden. Diese werden sich in der Regel für jedes Projekt bzw. jeden Einsatzzweck unterscheiden. Folgende Anforderungen wurden für epolos definiert:

- Einfache Bedienbarkeit
- Möglichkeit eigener Erweiterungen
- Anpassbarkeit des Layouts
- Gute Performance
- Deutsches Sprachpaket
- Unterstützung von e-Learning Standards (z.B. IMS¹⁵, SCORM¹⁶)
- Möglichst niedrige Kosten

Eine einfache Bedienbarkeit insbesondere für Lehrende wurde als besonders wichtig angesehen. Eine schwierige Bedienung würde aus epolos-Sicht zu einer verminderten Nutzung bzw. höheren Schulungskosten führen. Ebenso wichtig ist die Möglichkeit, eigene Erweiterungen am System durchführen zu können, da es kein LMS gibt, welches alle Anforderungen erfüllt. Dazu sind diese für den jeweiligen Einsatzzweck bzw. die Zielgruppen zu unterschiedlich. Da WebCT diese beiden Voraussetzungen bereits nicht erfüllte, schied es aus. Neben diesen beiden Anforderungen sollte das Layout anpassbar sein, um z.B. die Vorgaben hinsichtlich der Corporate Identity der Universität unterstützen zu

11 <http://www.blackboard.com/>

12 <http://www.im-c.de/>

13 <http://www.hyperwave.de/>

14 <http://www.webct.com/>

15 <http://www.imsproject.org/>

16 <http://www.adlnet.org/>

können. Wichtig für den Einsatz in Lehrveranstaltungen in Universitäten mit einer insgesamt zu erwartenden sehr hohen Nutzerzahl ist eine gute Performance und hohe Verfügbarkeit. Dies wird häufig von Open Source-Produkten nicht gewährleistet. Ebenfalls wichtig ist ein deutsches Sprachpaket, um den Einsatz des LMS nicht durch eine Sprachbarriere zu behindern. Um nicht dauerhaft an ein System gebunden zu sein, was insbesondere in Hinblick auf Lizenzpolitik und Kosten, aber auch auf veränderte Anforderungen wichtig ist, sollten e-Learning-Standards unterstützt werden. Diese Unterstützung ist ebenfalls bei Open Source-Systemen zu beachten, denn ein Wechsel von einem oder auf ein System, das keine Standards unterstützt, würde zusätzlichen Entwicklungsaufwand bedeuten und somit neue Kosten verursachen.

Nach Festlegung dieser Anforderungen erfolgte eine eingehende Untersuchung der oben erwähnten Lernmanagementsysteme durch Installation von Evaluationsversionen (soweit diese beschafft werden konnten), Informationsbesuche bei Referenzkunden sowie das Durcharbeiten von white papers, Produktdokumentationen oder technical reports. Am informativsten davon waren Informationsbesuche bei Referenzkunden. Insbesondere die white paper und technical reports sollten einer kritischen Betrachtung unterzogen werden, da diese häufig zu positiv dargestellt wurden.

Anschließend wurden die Erkenntnisse über die Systeme nach den genannten Kriterien bewertet und die Ergebnisse zur Auswahl eines Systems genutzt. Für epolos wurde Blackboard Learning System ML gewählt. Die ML-Version von Blackboard bietet verschiedene Sprachpakete, unter anderem auch ein deutsches. Daneben ist Blackboard ML eines der am einfachsten zu bedienenden Lernmanagementsysteme, das aber dennoch über eine umfangreiche Funktionalität verfügt. Es erfüllt damit das wichtigste Kriterium. Des Weiteren lässt sich das System über die so genannten Building Blocks¹⁷ erweitern. Es bietet eine gute Performance und unterstützt IMS und ansatzweise auch SCORM. Als negativ sind die hohen Kosten anzusehen, die sich aber durchaus auf demselben Niveau wie vergleichbare kommerzielle Systeme bewegen. Hinzu kommt, dass es sich in der ML-Version nicht gut anpassen lässt. Blackboard wird seit Oktober 2002 an den Universitäten Oldenburg und Osnabrück betrieben und in der Lehre eingesetzt.

Um einen erfolgreichen Einsatz zu gewährleisten, sind allerdings Serviceleistungen nötig, die im Folgenden beschrieben werden.

3.4 Einführung eines Lernmanagementsystems

Für den Einsatz im ersten Semester des Betriebes (Wintersemester 2002/2003) war die Nutzung in einigen wenigen Veranstaltungen vorgesehen, hauptsächlich

17 <http://buildingblocks.blackboard.com/>

im Bereich Informatik und Wirtschaftswissenschaften. Parallel dazu wurden Lehrende durch Präsentationen auf verschiedenen Ebenen geworben, um ihre Veranstaltungen ebenfalls durch ein Lernmanagementsystem zu unterstützen. Diese „Marketingmaßnahmen“ sind als wichtiger Punkt für den Erfolg einer Einführung anzusehen.

Als vielleicht wichtigste Dienstleistung ist daneben aber das Angebot von Schulungen zu sehen. Diese werden an der Universität Oldenburg durch das CELab vorgenommen. In der Regel wird jeder Lehrende, der seine Veranstaltung durch ein Lernmanagementsystem unterstützen möchte, individuell betreut. Dies lässt sich natürlich nur durchführen, solange die Zahl der Lehrenden, die ein solches System nutzen möchten, nicht zu groß wird. Durch diese Individualschulungen wird ein Multiplikatoreffekt erzeugt, der für einen erfolgreichen Einsatz sehr wichtig ist. Lehrende berichten im Kollegenkreis von ihrem Lernmanagement Einsatz und schaffen dadurch Anreize für andere Lehrende, dieses auch in ihren eigenen Veranstaltungen durchzuführen.

Des Weiteren sind Demokurse und Lernvideos sinnvoll, um mögliche Szenarien demonstrieren zu können und den Lehrenden entsprechende Hilfen zu geben. Hierzu gehört auch die Erstellung von Referenzmodellen, die die Unterstützung traditioneller Lehrveranstaltungen durch ein Lernmanagementsystem zeigen.

4 Referenzmodelle für traditionelle Lehrveranstaltungen

Unter einem Referenzmodell für traditionelle Lehrveranstaltungen soll hier im Gegensatz zu den Referenzmodellen in Rinn (2002) eine UML¹⁸-Darstellung einer universitären Lehrveranstaltung verstanden werden. In Abbildung 1 ist dies für ein Seminar dargestellt, wie es üblicherweise im Fachbereich Informatik durchgeführt wird.

Ein Lehrender verteilt zu Beginn Seminarthemen, die von den Studenten gewählt werden können. Anschließend bearbeiten die Studenten das Thema, erstellen eine Seminararbeit und halten dazu einen Vortrag. Der Lehrende bewertet diese Seminararbeit und der Student erhält für Vortrag und Arbeit eine Note. Dies ist im UML-Anwendungsfalldiagramm (s. Abbildung 1) durch die Anwendungsfälle „holding a seminar“, „providing subjects“, „literature research“, „participating in a seminar“, „writing seminar paper“, „preparation of lecture“, „supervision“ und „assessment“ dargestellt.

Die Kästen, die die Anwendungsfälle umschließen, zeigen die Unterstützung durch ein Lernmanagementsystem. Dies ist für alle Lernmanagementsysteme

18 <http://www.uml.org/> Die *Unified Modeling Language*TM (UMLTM) ist eine durch die Object Management Group (OMG) standardisierte graphische Sprache zur Beschreibung objektorientierter Modelle.

nahezu identisch, solange sie die entsprechende Funktion unterstützen, mit der ein Kasten bezeichnet ist. Insbesondere der Kasten „Building Block“ ist allerdings spezifisch für das Blackboard-System. Es handelt sich dabei, wie bereits oben erwähnt, um eine Erweiterung des Systems. Aus der Grafik ist abzulesen, dass z.B. „holding a seminar“ durch die Kursverwaltung des Lernmanagementsystems unterstützt wird.

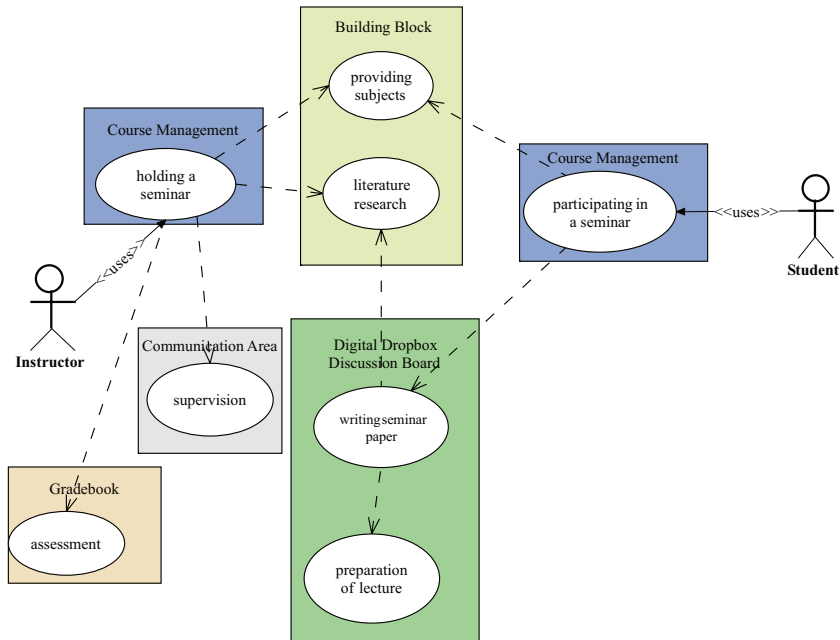


Abbildung 1: Referenzmodell eines Seminars

Mit Hilfe dieser Referenzmodelle, die für eine Vielzahl üblicher Lehrveranstaltungen entwickelt wurden und für die das Beispiel Seminar stellvertretend sein soll, wird Lehrenden grafisch dargestellt, mit welchen Mitteln des Lernmanagementsystems ihre Veranstaltung unterstützt werden soll. Abschließend soll eine kurze statistische Auswertung den bisherigen Erfolg dieser Maßnahmen belegen.

4.1 Statistische Auswertung

Die Nutzerstatistik (s. Abbildung 2) zeigt, dass es von Beginn des Einsatzes an einen starken Anstieg der Nutzerzahlen gegeben hat. Im April 2003 sind ca. 1900 Benutzer am System angemeldet. Davon ist ein Teil allerdings in keinem Kurs angemeldet und somit kein Teilnehmer. Hierbei handelt es sich häufig um Nutzer, die das System zunächst erproben möchten. Dieses ist durch die zweite obere Linie in der Grafik gekennzeichnet. Die untere Linie zeigt eine kontinuierlich ansteigende Zahl von Lehrenden, die das System nutzen. Diese Lehrenden betreiben derzeit etwa 80 Kurse und 40 Testkurse.

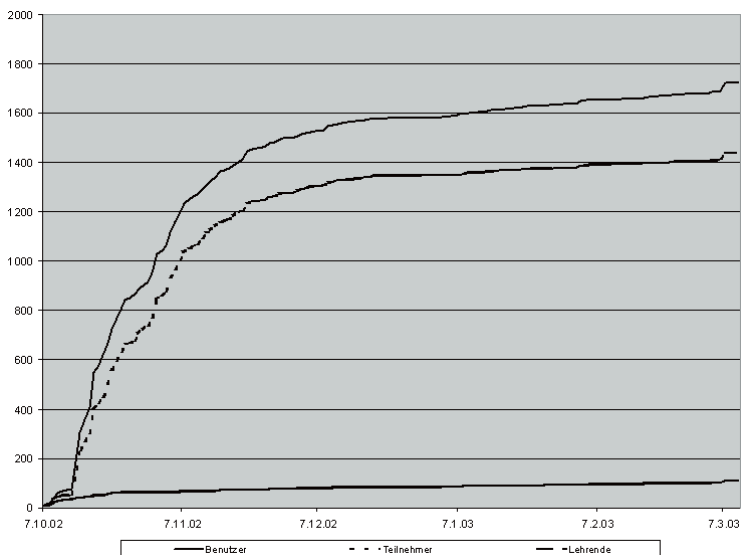


Abbildung 2: Nutzerstatistik des Blackboard-LMS an den Universitäten OL/OS

Parallel dazu wurden alle Veranstaltungen der Universität Oldenburg in das System importiert, um den Studierenden die Möglichkeit zu geben, auch nicht durch Lehrende unterstützte Veranstaltungen zu nutzen und Termine und Raumankündigungen im System zu erfahren. In einer Umfrage unter den Lehrenden wurden vor allem die Kommunikationstools als sehr wichtig angesehen, da sie den Studierenden einen Austausch untereinander ermöglichen und bei Schwierigkeiten und Problemen Hilfestellung durch andere Studierende bieten. Dies würde dann eine wesentliche Verbesserung gegenüber der Situation vor Einsatz eines Lernmanagementsystems bieten. Der Erfolg dieser Maßnahme muss sich aber erst noch zeigen.

Insgesamt kann also gesagt werden, dass die Maßnahmen zu einer erfolgreichen Nutzung und dem erfolgreichen Einsatz eines Lernmanagementsystems an den Universitäten Oldenburg und Osnabrück geführt haben.

5 Zusammenfassung und Ausblick

Dieser Beitrag gibt einen kurzen Überblick über den derzeitigen Stand (Mai 2003) der Einführung eines Lernmanagementsystems an den Universitäten Oldenburg und Osnabrück. Es wird ein Beispiel für einen generellen Auswahlprozess eines Lernmanagementsystems vorgestellt und gezeigt, wie und mit welchen Anforderungen ein solcher für den ELAN-Piloten epolos durchgeführt wurde. Weiter werden Maßnahmen dargestellt, mit deren Hilfe Lehrende beim Einsatz von e-Learning in der universitären Lehre unterstützt werden und wie Lehrenden durch Referenzmodelle traditioneller Lehrveranstaltungen Unterstützung beim Einsatz eines Lernmanagementsystems gegeben werden kann. Eine kurze statistische Auswertung zur Nutzung zeigt den Erfolg der Maßnahme.

Umfragen, die zur Evaluation unter den am Blackboard-System teilnehmenden Lehrenden und Studierenden durchgeführt wurden, haben bestätigt, dass das System überwiegend positiv angenommen wird. Es wurden allerdings auch einige Schwachstellen erkannt, die durch Erweiterungen des Lernmanagementsystems ausgeglichen werden sollen. Hierzu gehören unter anderem ein wesentlich verbessertes und auf Anforderungen der Lehre zugeschnittenes Diskussionsforum sowie ein internes Mailsystem. Diese Komponenten werden derzeit entwickelt.

Literatur

- [1] P. Baumgartner, H. Häfele und K. Maier-Häfele, E-Learning Praxishandbuch: Auswahl von Lernplattformen Marktübersicht - Funktionen - Fachbegriffe, 2002, Studien Verlag, Innsbruck
- [2] E.-E. Doberkat et al., Anforderungen an eine e-Learning-Plattform, 2002, Universität Dortmund and Universität Paderborn
- [3] U. Rinn, J. Wedekind (Hrsg.), Referenzmodelle netzbasierten Lehrens und Lernens, 2002, Waxmann Verlag, Münster
- [4] R. Schulmeister, Selektions- und Entscheidungskriterien für die Auswahl von Lernplattformen und Autorenwerkzeugen, 2000, Universität Hamburg IZHD

- [5] B. G. Wilson, Trends and Futures of Education: Implications for Distance Education, 2001, Quarterly Review of Distance Education Oktober 2001 (online Journal)

Kontakt

Dennis Reil
OFFIS
26121 Oldenburg
Dennis.Reil@offis.de

Reflections on the Value Chain of Scientific Information

Hans E. Roosendaal, Peter A. Th. M. Geurts¹

University of Twente, The Netherlands

Justification

This paper is a record of an invited presentation given at the IuK conference held in March 2003 in Osnabrück. As such the paper cites freely from previous publications of the authors, albeit sometimes in a different context resulting in some different and new observations and conclusions. The reader will find these specifically in chapter 4.2. The used citations are clearly marked in the text of this paper.

Abstract

This paper discusses recent and possible future developments in the value chain of scientific information under the influence of engines for change, such as the use of ICT and developments in research and education.

The increasing use of ICT in information for research and education will further the creation of institutional repositories at individual academic institutions or at consortia of academic institutions. These repositories will store the institution's research information production in a way that is sustainable and lends itself for direct distribution to other institutions around the world. These developments have a major impact on the positions of all stakeholders in the value chain. The establishment of institutional repositories by academic institutions may well lead to a structural change from a high margin/low volume to a low margin/high volume scientific information market.

It is argued that such a new value chain in scientific information is feasible, in particular if the international developments in education are taken into the equation and are fully grasped by the academic institutions. Such developments are the introduction of the Bachelor-Master system in academic education through-

¹ Hans Roosendaal is professor of scientific information in the School of Business, Public Administration and Technology, and in the Department of Computer Science. Peter Geurts is methodologist and works at the School for Business, Public Administration and Technology. This research is carried out in the Centre for Telematics and Information Technology (CTIT).

hout Europe and the new opportunities for distance and lifelong learning that ICT provides.

1 Introduction

Stakeholder-based systems are characterised in their development by the optimising behaviour of these stakeholders. The scientific information is such a system and is governed by the following:

“Authors want to publish more, readers want to read less.”

This law states that wide exposure is paramount to the author of scientific information and that (pre)selection is paramount to its reader. It determines largely the value chain of scientific information. Any force like the use of Information and Communication Technology (ICT) by the stakeholders involved (authors, readers, academic libraries, scientific publishers etc.) that allows better fulfilling this law is an engine for change of the value chain. And this will prompt changes in the roles of the stakeholders in scientific communication.

1.1 Complex information and empowerment²

Scientific information is a particular example of a class of information called complex information. Complex information is essentially information that is integrated over the entire value chain of a particular process. Complex information is therefore per definition information of strategic relevance, be it for an entire industry or business, for an entire enterprise or for any larger organisation, be this a private or public organisation or even a private-public co-operation. A good example of such an organisation being dependent on complex information is an academic institution, or generally a knowledge-intensive organisation.

The information carried in each individual link of the value chain will be of a different nature, form, norm, standard etc than in another link in the chain. The intriguing issue then is to integrate the information carried by the various links in such a way as to create a fully integrated information chain over the entire value chain.

In such a value chain we can distinguish between generic stakeholders or main actors and institutional stakeholders or facilitators or intermediaries. Using this terminology we see that in the value chain for scientific information the author and the reader are the generic stakeholders and universities, publis-

2 Hans E. Roosendaal, Theo W.C. Huibers, Peter A. Th. M. Geurts and Paul E. van der Vet, “Changes in the value Chain of Scientific Information: Economic Consequences for Academic Institutions and their Information Provision”, Online Information Review, accepted for publication

hers and libraries and other facilitators are institutional stakeholders. An important engine for structural change in the value chain is the drive for empowerment of the generic stakeholders in relation to the institutional stakeholders. In particular ICT (Information and Communication Technology) is an important engine for change as it holds the promise of further empowerment of the generic stakeholders, and in particular of the author as the most intensive user of the system. ICT empowers the author to increase the visibility of his/her work and it helps the reader in digital searching of the information base. But, ICT is foremost a powerful engine for change in the value chain because it allows the transition from a low volume/high margin business, as is the current mode in the industry of research information, to a high volume/low margin business³. This strategic development drives the academic institution to introduce ICT-supported education and research widely. For the academic institution this development towards widely applying ICT in education is considered an autonomous development⁴. This desire to empowerment of the generic stakeholders articulates the need for a new value chain in scientific communication.

Other engines for change that are relevant today are the transition of the European universities to the Bachelor-Master system and the general developments in research and education. Main developments in research are the increasing emphasis on strategic research⁵ and the role of intellectual property. In education the promises of distance and lifelong learning in connection with ICT play an important role. The developments in the value chain will be discussed in detail in chapter 3, after a brief discussion of the main functions in scientific communication in the next chapter. Strategic issues resulting from the developments in the value chain are discussed in chapter 4.

2 Functions in scientific communication

In this paper we limit ourselves to a brief discussion of the main functions of scientific communication and refer the reader to previous publications for a full discussion.

3 See e.g.: Nicholas Negroponte, "Being Digital", Hodder and Stoughton, London, 1995

4 See e.g.: John Seeley Brown and Paul Duguid, "Universities in the Digital Age", Heldref Publications, Washington, 1996

5 M. Gibbons, „The new production of knowledge: The dynamics of science and research in contemporary societies“, Sage Publications (1994).

Following^{6, 7, 8} there are four main functions of scientific communication: the registration, awareness, certification and archiving functions (see Figure 1).

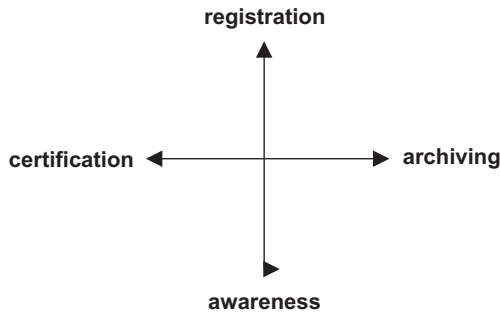


Figure 1: Functions in scientific communication

Figure 1 shows the overall communication process, as it is part of the research and education process. It also shows which parts of the communication process can be externalised out of the research and education process.

Following the familiar classification of the functions into author and reader functions⁹ we see that the functions of registration and certification are author functions and the functions of archiving and awareness are reader functions. Functions can be considered internal or external to the research and education process and functions. By their nature, registration and archiving as external functions to the research and education process can be easily outsourced to external stakeholders, viz. the publisher and the library. The internal functions are an integral part of the research and education process.

For a comprehensive discussion of these functions and the methodological choices and consequences the reader is referred to references. In this paper it

6 Hans E. Roosendaal and Peter A. Th. M. Geurts, „Forces and Functions in Scientific Communication: an Analysis of their Interplay“, Proceedings of the Conference on „Co-operative Research in Information Systems in Physics“, University of Oldenburg, Germany, September 1-3, 1997.

<http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>

<http://www.ub.utwente.nl/webdocs/dinkel/1/00000001.pdf>

7 Hans E. Roosendaal, Peter A. Th. M. Geurts and Paul E. van der Vet, “Developments in scientific communication: Considerations on the value chain”. Information Services and Use, 21 (2001) 13-32

8 Hans E. Roosendaal, Peter A. Th. M. Geurts and Paul E. van der Vet, “Developments in scientific communication - The virtual marketplace as a prerequisite for growth”, “A Century of Science Publishing” edited by E.H. Frederiksson, IOS Press, Amsterdam, 2001

9 H.E. Roosendaal, „Scientific Communication and the Science Process“, Proceedings of the International Summer School on the Digital Library, Tilburg, The Netherlands, August 1996, ed. J.G. Van Luyt-Prinsen and E. Meijer, Ticer B.V. (1996), 1-12.

suffices to conclude that the use of the four functions provides a consistent analysis of formal and informal scientific communication.

3 Value chain

3.1 General remarks

The value chain is defined as the chain of links of values added by each stakeholder in the entire process of scientific information. The value chain defines the mutual roles of these stakeholders and their responsibilities. The original idea of the value chain as presented by Porter is based on the process view of organisations, the idea of seeing a manufacturing or service organisation as a system made up of subsystems each with inputs, transformation processes and output. Inputs, transformation processes and outputs involve the acquisition and consumption of resources¹⁰.

The links of added value in the value chain can always be mapped on the four main functions of scientific communication: the registration, awareness, certification and archive functions.

The value chain then starts with the creation of the work by the author who in the traditional value chain (see figure 1) submits this work to a journal or other publication outlet at a publisher, for registration. If this is a “for profit” or “not for profit” publisher is irrelevant, it is the publishing function that is being addressed. The publisher sends the paper out for review, for certification. Possibly after some rounds of revisions, the publication is accepted for publication resulting in the registration of the work. The publication will then be distributed, often with the subscription agent as an intermediary between the publisher and the university. The publication then lands at the academic institution, in the institution’s library to be archived and finally on the desk of the reader who will become aware of the publication.

The above value chain is the familiar traditional value chain for research information. However, looking at educational information we do see a very similar chain. Again, the information is being created by the author, be it the teacher or the student, and submitted for registration. The information is being certified, either by the teacher in case of student material, by a publisher in case of course material, or by an accreditation board in case of an entire course. The material will be distributed either internally within the institution, or it may be shared with other institutions, directly or indirectly via a publisher. In the end it will land with the reader, be this a teacher or student.

10 Michael E. Porter, “Competitive Strategy”, The Free Press, New York, 1980.



Figure 2: traditional value chain

3.2 Some changes in the value chain due to ICT?

In general, some major changes in the value chain due to the introduction of ICT have been proposed,¹¹

It has been proposed to reduce the value chain to the author, reviewer and reader only, i.e. essentially without any intermediaries from outside the research process, be it the publisher, the agent, the academic institution or the library of this institution. The author will be able to self-distribute the work via the Web and any reviewer can pick it up to add his comments. Or even the reviewer can be missed entirely as the reader may well be qualified to judge the work by himself. However, there are a number of issues at stake: how is the marketing of the work done, how will the reader find the publication, and how can publications be filtered? And there is the problem of certification: who will select the reviewer? Evidently the author should not and the reader cannot do this.

Another possibility will be to take the academic institution out of the chain. In a digital environment it may well be feasible that the publisher is responsible for the entire communication between author and reader. The only problem is that the publisher will be responsible for the archive. It is probably not in the publishers' interest to assume this responsibility and it is very likely that the academic community will not entrust this responsibility to the publisher. Alternatively, we can replace the publishers entirely by the academic institution at both ends of the chain, at the author end and the reader end. But, who will select the referee? This cannot be the task of the institution of the author. This will lead academic institutions to create a wider forum in order to be able to better accomplish this function by entering into alliances with other academic institutions. In the end, this will most likely lead to a professionalisation of the publishing function within this alliance: in other words, the "publisher" will be reinvented. This is indeed a direction that some academic institutions and alliances have taken by (re)establishing university presses. The result will be that more, new journals and other information products will be added to the already existing product range of publishers furthering the further proliferation of research journals and other information products.

11 See e.g.: Don Tapscott, "The Digital Economy: Promise and Peril in the Age of Networked Intelligence", The MacGraw-Hill-Companies Inc., New York, 1996

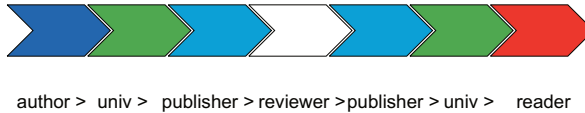


Figure 3: emerging value chain

The need for certification of research information requires a partner independent of the institutions. This is a role for a “sort of publisher”. This may lead to the following alternative value chain: both the academic institution and the publisher will have a symmetrical position with respect to the key actors in the chain, the author and the reader. The institution is in direct contact with the author and the reader and will take responsibility for the necessary disclosure of the information and the required technology. The publisher is responsible for the dissemination and branding and assists editorial boards in organising the certification process. See Figure 2.

An important consideration is that the academic institution is in the position to take up this role, as ICT will allow the institution to create new options for distance learning and continuing professional education or even to innovate the education system at large. As we have seen that the value chain for educational information is very similar indeed, it will not be surprising that we can envisage similar changes in the value chain arising from the innovation of education triggered by the introduction of ICT in education. In order to be successful in this innovation process the institution will be required to create an ICT-supported publishing and archiving environment for the entire information process.

A consequence of the above observations is that we cannot separate the developments in the value chain for the library, and their economic consequences, from the developments in the value chain for the parent institution or even the entire scientific information infrastructure.

4 Strategic issues

4.1 General considerations

Widespread use of ICT allows academic institutions to change from supply-push to demand-pull education. This will open up these institutions for new cohorts of students, in particular virtual “Wanderstudenten”, leading to new opportunities for academic institutions.

Developments in distance learning and continuing professional development lead to effectively expanding the time window that students will dwell at the academic institution by an estimated factor of two, i.e. from 5-7 years to something

like effectively 10-15 years per student-lifetime. The transformation of the educational system towards a system based on customisation rather than massification is designed to lead to a higher mobility of students within the international research and educational system world-wide. It is amongst others for this reason that Europe-wide a new bachelor-master-PhD system is being introduced. This development will have severe consequences for the financing and rewarding systems of the academic institutions and will open up novel options for public-private collaboration between these institutions and relevant industry world-wide. Next to collaboration within academia, collaboration between academia and industry will become a normal mode. No doubt, this will have a severe impact on the economic conditions under which the academic institution will have to operate in future and the benefits for the institution of such a new economic regime will crucially depend on the institution's investments in a publishing and archiving environment. This is the very reason for a fundamental change in the value chain as it represents a substantial change in the position of the academic institution in supporting and representing the main actors in the value chain: the authors and the readers, now being students, teachers as well as researchers. These actors will make use of a variety of overlapping information sources depending on their specific needs at a given time. It will support the trend that students will start as junior researchers gradually becoming more senior as they progress in their student career. Such a development to a conceptual integration of research and education is commensurate with the bachelor-master-PhD system now being introduced throughout Europe. A prerequisite for such a development is adequate access and navigation for the student over the various information sources. Adequate access to and navigation in scientific information, i.e. both for research and education, turns out to be key for further conceptual integration of research and education, a universal goal of all times. It is the application of ICT that makes this goal attainable.

Achieving this goal requires a degree of technical integration of scientific information ranging from interoperability between the various sources at one institution to a federated system between institutions and even to a fully technically integrated system. For the discussion of this paper it suffices to note, as we have already noted above in discussing the value chain, that both research information and educational information are subject to very similar requirements, in particular with respect to the registration, archiving and certification function, albeit that the actors may be different. The desired mobility of the actors, in particular of the most mobile actor the student, requires an actor centred information system for the academic institution.

The above leads to an academic institution-supported mode of “self publishing and archiving” (see also Harnad¹²) for research and education. To establish such an environment requires the application of the same technologies as for research information only. A quick estimate shows that the total volume of production of educational information per institution would surpass by about an order of magnitude its required production of research information only. This means that the institution is in the position to piggy-back the production, archiving, distribution and dissemination and of research information on the production, archiving, distribution and dissemination of their educational information. Albeit, that for archiving it may be noted that archiving for education is in general shorter term archiving than archiving for research.

In summary, academic institutions will have to and will develop their own internal value chain for information for research and teaching in order to prepare themselves for a future of open exchange of this information in a new regime of business models, or in a new value chain. It may be noted that the word business model is being used throughout this article as a model for the sharing, exchanging and trading of information between stakeholders of possibly different nature, non-commercial or commercial and all possible permutations.

4.2 Strategic relations in the value chain

Following the above we start from the premise that the creation by academic institutions of a publishing and archiving environment integrating the information for research and teaching is a key strategic issue in scientific communication. This means that academic institutions will each build institutional repositories for research and teaching. These institutional repositories will allow a global network of repositories. Next to these institutional repositories carrying the works of the employees of the academic institution such a network also has to cater for certification of the material. This requires the international organisation of entities independent of the academic institutions, as we have seen when discussing the value chain in chapter 3.2.

Furthermore, institutional repositories will allow academic institutions to share and trade scientific information between them using a variety of appropriate business models of choice within academia and between academia and private industry. A clear vision on the value chain of scientific information is a prerequisite for developing adequate business models.

This no doubt calls for a publishing role in such a network. This publishing role can be organised between and by the academic institutions, or by other organisations of a public or private character or by publishers, be these ‘for profit’ or ‘not for profit’ publishers. This publishing function is a facilitating and ag-

12 Stevan Harnad, “The self-archiving initiative”,
<http://www.nature.com/nature/debates/e-access/Articles/harnad.html>

gregating role between the repositories by providing access and linking to research and teaching information stored on the various different sources of creation or institutional repositories. This function can and probably will introduce a quality filter as added value by giving access to specific certified material thereby introducing editorial control on the material. It seems obvious that for research information but probably also for teaching information such a certification function should carry an international dimension.

As stated above this function can be assumed by a number of new or existing public or private institutions or by the publishers. Each of these stakeholders has to prove the value it adds to the information chain in terms of the functions that it provides. It would mean that a number of the existing stakeholders will have to assume new roles and will have to prove themselves in such new roles in the new and emerging value chain. The academic institutions and therefore the generic stakeholders, the author and the reader, will have an important position in this new value chain. In the end it will be the generic stakeholders, and in particular the author as the author has the first right of choice, who will decide on the efficacy and efficiency of the scientific communication network in his/her desire for empowerment. The value chain represents the ultimate competitive landscape, it is not an instrument for public regulation.

The above sketched development means that we are gradually moving towards a global, distributed network of scientific communication as a network of repositories, archives, libraries, academic institutions and facilitators and aggregators. Such an international and distributed network demands an international and distributed ownership, or a strategic management involving all the stakeholders, generic as well as institutional, involved in the value chain, albeit with different weights commensurate to the values added and the functions performed in the network. The primary task of this kind of strategic management is to facilitate all stakeholders in realising their individual strategic and operational goals. Needless to state that these individual stakeholder goals can and will be competing. Such a development then also calls for new trust relations¹³ between the stakeholders for such a new mix of competition and co-operation in the construction of a global scientific information infrastructure that is not needed in an infrastructure for the distribution of paper-based information.

Establishing new trust relations is not a sufficient task, it certainly is a necessary task. It is a prerequisite for creating new business models between academic institutions and other stakeholders, and new distribution models on the basis of such new business models.

13 Harry Hummels and Hans E. Roosendaal, "Trust in Scientific Publishing". *Journal of Business Ethics*, 34 (2001) 87-100

5 Concluding remarks

This paper discusses some developments in the value chain of scientific information. Most important engines for change are the widespread use of ICT coupled with developments in research and education. Examples are the introduction of the Bachelor-Master system in academic education throughout Europe and the new opportunities for distance and lifelong learning.

The increasing use of ICT in information for research and education will lead to the creation of institutional repositories at individual academic institutions or at consortia of academic institutions. This means that the institution's research information production will be stored on these repositories in a way that is sustainable and leans itself for direct distribution to other institutions around the world. This development clearly marks a strengthening in the position of the generic stakeholders striving for empowerment, the author and the reader, and the academic institutions in the value chain. And this again will induce new roles and positions for the other institutional stakeholders such as publishers and other intermediaries. Due to this shift in roles the competition between these institutional stakeholders will be much more focused on the specific value added and the functions performed by these stakeholders. The establishment of institutional repositories by academic institutions will create an option to change from a high margin/low volume to a low margin/high volume scientific information market.

It is argued that this new value chain in scientific information is feasible, in particular if the international developments in education are taken into the equation and are fully grasped by the academic institutions.

SozioNet - Web-Ressourcen für ein künftiges Fachportal Sozialwissenschaften

Natascha Schumann, Wolfgang Meier, Sue Heise, Rudi Schmiede

Technische Universität Darmstadt, Institut für Soziologie

Abstract

SozioNet is part of a forthcoming social science information portal, which is currently being developed by the infoconnex initiative. SozioNet provides access to freely available resources with relevance to social sciences. SozioNet aims at building a network of social science insitutions and scientists, to agree on and establish common metadata standards. SozioNet implements a general infrastructure for the creation of semantically rich metadata and for harvesting and retrieveaing relevant resources in a domain specific focus.

1 Einleitung

Infoconnex¹ ist eine Initiative, welches einen integrierten Zugang zu unterschiedlichen Informationsangeboten aus den führenden Fachdatenbanken aus Pädagogik, Sozialwissenschaften und Psychologie ermöglicht. In der Regel handelt es sich hierbei um kostenpflichtige Ressourcen, aber es wird auch sog. „Graue Literatur“ zum kostenlosen Zugriff bereitgestellt werden. SozioNet ist ein Bestandteil des im Rahmen von infoconnex entstehenden Fachportals Sozialwissenschaften.

SozioNet versteht sich als zusätzliches Angebot zu bereits bestehenden Informationsdienstleistungen im sozialwissenschaftlichen Bereich. Dazu gehören die verschiedenen Fachdatenbanken und Angebote des IZ Sozialwissenschaften, die Literaturdatenbank SOLIS², die Forschungsdatenbank FORIS³, die Institutionendatenbank SOFO⁴ und das Clearinghouse SocioGuide⁵. Ein weiteres Angebot ist die Virtuelle Fachbibliothek Sozialwissenschaften ViBSoz⁶, die Zu-

1 <http://www.infoconnex.de>

2 <http://www.gesis.org/Information/SOLIS/>

3 <http://www.gesis.org/Information/FORIS/index.htm>

4 <http://www.gesis.org/Information/SOFO/index.htm>

5 <http://www.gesis.org/SocioGuide/index.htm>

6 <http://www.vibsoz.de/>

gang zu unterschiedlichen Beständen sozialwissenschaftlicher Literatur ermöglicht,

Das Ziel von SozioNet besteht in der Vernetzung frei zugänglicher Internet-Ressourcen. Die Idee dazu besteht bereits seit 1998; seitdem haben eine Reihe vorbereitender Treffen stattgefunden. Seit April 2002 wird das Projekt SozioNet vom Bundesministerium für Forschung und Bildung (BMBF) im Rahmen des Verbundprojekts infoconnex gefördert.

SozioNet bietet einen bequemen Zugang zu Ressourcen, die verteilt auf den Servern der jeweiligen Einrichtungen vorhanden sind. Darüber hinaus möchte SozioNet einen Beitrag dazu leisten, die Sichtbarkeit sozialwissenschaftlicher Einrichtungen und ihrer Aktivitäten zu verbessern. Außerdem sollen Sozialwissenschaftlerinnen und Sozialwissenschaftler motiviert und darin unterstützt werden, ihre Publikationen über das Internet zu verbreiten und mit standardisierten Metadaten auszuzeichnen.

SozioNet wurde stark beeinflusst durch Projekte wie MathNet⁷, PhysNet⁸ und SOSIG⁹ (Social Science Information Gateway). Allerdings unterscheidet sich SozioNet in einigen Punkten von diesen Vorbildern, was die zugrunde liegende Technik und den Gebrauch von Ontologien und Standards, wie beispielsweise OWL¹⁰ (Web Ontology Language) betrifft.

Bislang beteiligen sich elf sozialwissenschaftliche Einrichtungen an SozioNet. Sie werden ihre Ressourcen mit Metadaten auszeichnen, Schnittstellen zu bereits in Datenbanksystemen vorgehaltenen Daten herstellen und die von uns bereitgestellten Metadatenschemata und Werkzeuge implementieren, evaluieren und weiterentwickeln.

Langfristig wird eine Etablierung von SozioNet innerhalb der Sozialwissenschaften angestrebt. Dafür ist es notwendig, mit den beteiligten Institutionen Organisationsstrukturen zu schaffen, um eine nachhaltige Nutzung und Erweiterung des Dienstes über die Projektlaufzeit hinaus zu gewährleisten. Hierzu bestehen entsprechende Vereinbarungen mit dem Informationszentrum Sozialwissenschaften.

Bislang besteht in den Sozialwissenschaften keine lebendige und etablierte Preprintkultur wie beispielsweise in den Naturwissenschaften. Jedoch wird das Internet auch in diesem Bereich zunehmend dazu genutzt, um unterschiedliche sozialwissenschaftliche Ressourcen zugänglich zu machen. Inzwischen stellen viele Einrichtungen die unterschiedlichsten Arten von Publikationen und Materialien zur Verfügung: Arbeitsberichte, Lehrmaterialien, Präsentationen, Artikel, Aufsätze, Projektbeschreibungen, Personen- und Institutionenseiten uvm.

7 <http://www.math-net.de/>

8 <http://physnet.uni-oldenburg.de/PhysNet/>

9 <http://www.sosig.ac.uk/>

10 <http://www.w3.org/TR/owl-guide/>

Diese liegen dann auf den Servern der jeweiligen Person oder Institution, so dass sie für Außen stehende nur schwer auffindbar sind.

2 Zielsetzungen

Das Ziel von SozioNet ist die Vernetzung frei zugänglicher Ressourcen von Wissenschaftlerinnen und Wissenschaftlern sowie von wissenschaftlichen Einrichtungen aus dem Bereich der Sozialwissenschaften im Internet.

Fachspezifische wissenschaftliche Ressourcen sind über herkömmliche Suchmaschinen nur schwer zu finden, da diese in der Regel nach algorithmischen Methoden (Vektormodell, prohabilitische Modelle etc.)¹¹ vorgehen und den fachlichen Kontext des Dokuments vernachlässigen. Die Auszeichnung von Ressourcen mit standardisierten Metadaten ermöglicht ein fachspezifisches Retrieval, das über die rein wortbezogene Suche hinausgeht und gezielt sozialwissenschaftliche Internetquellen auffindet.

Mit Hilfe von Metadaten kann bei der Suche der fachliche Kontext berücksichtigt und somit die Relevanz einer Ressource einfacher und schneller beurteilt werden. Ein aufwendiges Sortieren in endlosen Trefferlisten wird so umgangen.

Außerdem sollen interessierte Sozialwissenschaftlerinnen und Sozialwissenschaftler dazu animiert werden, ihre Publikationen mit Metadaten auszuzeichnen und verstärkt über das Netz zu vertreiben. Es soll innerhalb der sozialwissenschaftlichen Community eine überregionale Informationsinfrastruktur geschaffen werden, die von den Wissenschaftlerinnen und Wissenschaftlern selbst gestaltet und organisiert wird. Die Ressourcen selbst verbleiben in der Verantwortung und Pflege der veröffentlichenden Einrichtung oder Person, SozioNet organisiert lediglich den Zugang zu den Informationen und stellt Werkzeuge für die Metadatenerzeugung zur Verfügung.

Für ein erfolgreiches Retrieval ist es notwendig, sich auf einen gemeinsamen Metadatenstandard zu einigen, den alle benutzen und der möglichst viele für den Kontext von SozioNet relevante Aspekte berücksichtigt. Darüber hinaus sollen die verwendeten Metadaten internationalen Standards genügen, um eine Nutzung unabhängig von SozioNet und die Anschlussfähigkeit an zukünftige Entwicklungen zu gewährleisten.

11 Vgl. Salton, G. und M.J. McGill: *Introduction to Modern Information Retrieval*, New York 1983, und Baeza-Yates, R. und B. Ribeiro-Neto: *Modern Information Retrieval*, Harlow 1999

3 Metadaten

Innerhalb von SozioNet kommen Metadaten in zwei Bereichen zur Anwendung. Zum einen werden die vorhandenen Online-Ressourcen mit Metadaten ausgezeichnet, und zum anderen enthält die SozioNet-Seite Metadaten.

Für die Auszeichnung von sozialwissenschaftlich relevanten Online-Ressourcen wurde für SozioNet ein eigenes Metadatenschema (auch Ontologie genannt) entwickelt. Das Metadatenschema beschreibt die verschiedenen Arten von Ressourcen, z.B. „research paper“, „presentation“ etc., mit unterschiedlichen Eigenschaften.

Verwendet werden dafür in erster Linie Dublin Core Metadaten und ihre Qualifier¹². Als etablierter Standard dient Dublin Core der Erschließung von Internet-Ressourcen. 15 Kernelemente und mehr als 30 Qualifier dienen der Beschreibung. Da Dublin Core sich in erster Linie auf bibliografische Angaben beschränkt, werden nicht alle fachspezifischen Aspekte abgedeckt. Für die inhaltliche Erschließung werden daher zusätzlich der Thesaurus und die Klassifikation des IZ Sozialwissenschaften herangezogen. Diese liegen jeweils in RDF vor.

RDF¹³ bedeutet Resource Description Framework und ist eine Beschreibungssprache für Metadaten. Man unterscheidet zwischen RDF Modell und Syntax und RDF Schema. Das RDF Modell zerlegt jede Aussage in drei Teile, so genannte Triple. Das bedeutet, eine Ressource hat eine Eigenschaft und einen Eigenschaftswert. In RDF wird jede Aussage als Triple dargestellt. RDF Syntax ist die Repräsentation von RDF in XML. Im RDF Schema (RDFS)¹⁴ werden die grundlegenden Strukturen wie Klassen und Eigenschaften festgelegt. Es bietet die Möglichkeit der modularen Definition von Metadaten, d.h. es können bereits bestehende und eigene Schemata eingebettet bzw. modelliert werden. RDF Schema bietet jedoch nur begrenzte Ausdrucksmöglichkeiten in Bezug auf Klassenbildung und Beziehungen.

Als zusätzliche Abstraktionsschicht wird OWL¹⁵ (Web Ontology Language) verwendet. Diese Sprache bietet zusätzliche Modellierungsmöglichkeiten und ist etwas einfacher in der Handhabung als RDF. Die Klassifikation und der Thesaurus Sozialwissenschaften liegen in OWL vor.

Das SozioNet-Metadatenschema ist eine Ontologie und definiert verschiedene Objektklassen, wie beispielsweise Arbeitsberichte, Unterrichtsmaterialien, Diplomarbeiten etc. Außerdem werden die Beziehungen zwischen den unterschiedlichen Objekttypen festgeschrieben und Referenzen zum Thesaurus und der Klassifikation Sozialwissenschaften hergestellt. Die SozioNet-Ontologie

12 Eine Liste der Elemente und Qualifier findet sich unter: <http://dublincore.org/documents/dcmi-terms/>

13 <http://www.w3.org/TR/REC-rdf-syntax/>

14 <http://www.w3.org/TR/rdf-schema/>

15 <http://www.w3.org/TR/owl-features/> OWL ersetzt DAML (DARPA Markup Language)

definiert somit ein gemeinsames Vokabular für die Sozialwissenschaften, welches neben bereits existierenden Schemata wie beispielsweise Dublin Core auch domänenspezifische Konzepte beinhaltet. Das Schema ist im SozioNet-Namespace¹⁶ definiert.

Durch die Verwendung von aktuellen Standards ist die Nutzung der erzeugten Metadaten nicht auf SozioNet beschränkt; die Metadatensätze können beliebig genutzt werden.

Die erstellten Metadatensätze sollen möglichst auf den Servern der Einrichtungen verbleiben. Sie werden als XML-Datei in das gleiche Verzeichnis wie die Ressource selbst gespeichert. Langfristig soll dieser Ansatz der verteilten Systeme alleine bestehen, derzeit werden die Metadatensätze noch zusätzlich in einer XML-Datenbank¹⁷ vorgehalten.

Ein Beispieldatensatz macht diese verschiedenen Bestandteile des Metadatenschemas sichtbar:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/terms/"
  xmlns:sn="http://www.sozionet.org/1.0/#">

  <sn:ResearchPaperrdf:about="http://www.zeitschriftarbeit.de/
    docs/2-2000/wolf.PDF">
    <dc:title>Das Netzwerk als Signatur der Epoche</dc:title>
    <dcq:alternative>Anmerkungen zu einigen neueren Beiträgen zur
      soziologischen Gegenwartsdiagnose</dcq:alternative>
    <dc:creator>Wolf, Harald</dc:creator>
    <dc:subject rdf:resource="http://sozionet.org/1.0/
      topics#10220"/>
    >dc:subject rdf:resource="http://sozionet.org/1.0/
      topics#1080404"/>
    <dc:subject>Netzwerkgesellschaft</dc:subject>
    <dc:subject>Netzwerke</dc:subject>
    <dc:subject>Netzwerkforschung</dc:subject>
    <dc:subject>Organisationssoziologie</dc:subject>
    <dc:subject>Soziologische Theorie</dc:subject>
    <dcq:abstract>Der Aufsatz beinhaltet Aspekte der Netzwerkde-
      batte, die für eine soziologische Deutung der Gegenwarts-
```

16 <http://www.sozionet.org/1.0/#>

17 <http://exist-db.org>


```

gesellschaft empirisch wie theoretisch wichtig sind....
</dcq:abstract>
<dc:publisher> Landesinstitut Sozialforschungsstelle
Dortmund </dc:publisher>
<dcq:created>1999-07-14</dcq:created>
<dcq:modified>2000</dcq:modified>
<dcq:IMT>text/pdf</dcq:IMT>
<dc:language>de</dc:language>
<dcq:isPartOf rdf:resource="http://www.zeitschriftarbeit.de/">
<dcq:isFormatOf>Zeitschrift Arbeit, Heft 2, Jg.9 (2000),
S.95-104 </dcq:isFormatOf>
</sn:ResearchPaper>
</rdf:RDF>

```

Für die Erstellung eines solchen RDF-Metadatensatzes steht ein webbasiertes Formular zur Verfügung.¹⁸

4 SozioNet-Seite

Neben den Online-Ressourcen sollen auch institutionelle Informationen über SozioNet recherchierbar sein. Die meisten Einrichtungen halten solche Informationen vor, allerdings ist es sehr schwer, sich einen Überblick über die verschiedenen Institute hinweg zu verschaffen. Außerdem sind diese Informationen in jeder Einrichtung unterschiedlich aufbereitet und an unterschiedlichen Orten zu finden, so dass es für externe Besucher der Seite schwer ist, an die gewünschten Informationen zu gelangen.

Angelehnt an das Konzept der „secondary homepage“ oder auch „MathNet-Seite“¹⁹ aus MathNet soll die SozioNet-Seite eine Navigationshilfe für die Nutzerinnen und Nutzer sein, um gezielt zu institutsbezogenen Informationen zu gelangen. Die SozioNet-Seite ist in Struktur und Layout einheitlich für jedes Institut. Die Nutzerinnen und Nutzer können so ohne Kenntnisse der Strukturen der jeweiligen Institutsserver problemlos nach den gewünschten Informationen recherchieren.

Zunächst wird diese Seite Links zu den Seiten mit der gewünschten Information enthalten, z.B. über die Mitarbeiterinnen und Mitarbeiter. Sukzessive soll das Metadatenschema mit Objektklassen zu dieser Seite erweitert werden, so dass ein semantisch reichhaltiges Netz von relevanten Informationen entsteht. Das bedeutet, auch zu Personendaten oder Organisationen werden Metadaten

¹⁸ Vgl. Kapitel 6

¹⁹ <http://www.math-net.org/navigator/>

erfasst. So wird ein Überblick über die Aktivitäten innerhalb der Sozialwissenschaften ermöglicht, der einen echten Mehrwert für die Nutzerinnen und Nutzer erbringt.

5 Harvesting

SozioNet basiert auf der Selbstorganisation der Wissenschaftlerinnen und Wissenschaftler. Das bedeutet, dass die Ressourcen auf den jeweiligen Institutsservern verbleiben und die Verantwortung und Pflege weiterhin den Autorinnen und Autoren bzw. ihren Institutionen obliegt. Daher müssen die Metadaten der jeweiligen Dokumente von einem Harvester eingesammelt werden.

Zum Harvesting von Webressourcen stehen einige Software Tools zur Verfügung, beispielsweise die Open Source Harvest Software²⁰. Diese wird von vielen Projekten genutzt, zum Beispiel von MathNet und PhysNet. Harvest unterscheidet zwischen Gatherern und Brokern. Der Gatherer durchsucht je nach Vorgaben die Webseiten und sammelt Metadaten und/oder weitere Daten ein. Der Broker greift auf die vom Gatherer gesammelten Daten zu und leitet sie an den Nutzer weiter.

Das ursprüngliche Harvesting-Konzept bietet nicht die optimale Unterstützung für SozioNet, unter anderem deshalb, weil die Metadaten im SOIF²¹-Format ausgegeben werden. SOIF ist ein einfaches hierarchisches Format, während RDF eine komplexe Graphenstruktur unterstützt. Bei der Konvertierung der Metadaten von RDF nach SOIF besteht somit die Gefahr eines gewissen Informationsverlustes.

Aus diesem Grunde wurde in SozioNet eine gemischte Lösung entwickelt. Bei Einrichtungen, die ihre Materialien in internen Content-Management-Systemen vorhalten, werden die Metadaten über das OAI-MHP²² (Open Archives Initiative Metadata Harvesting Protocol) geholt. Die OAI²³ wurde 1999 mit dem Ziel gegründet, wissenschaftliche Publikationen möglichst schnell anderen Wissenschaftlerinnen und Wissenschaftlern zugänglich zu machen. OAI-MHP ist ein anwendungsunabhängiges, standardisiertes Protokoll zum Austausch von Metadaten. Die Einrichtung wird dadurch zu einem Dataprovider und SozioNet zu einem Serviceprovider, der die Metadaten abfragt. In diesem Fall muss der SozioNet-Harvester dann nur noch die Volltextindexierung übernehmen.

Der SozioNet-Harvester basiert auf verschiedenen frei zugänglichen Open Source Tools, unterstützt von einem einfachen Komponentenmodell. Der Har-

20 <http://harvest.sourceforge.net/>

21 SummaryObject Interchange Format

22 <http://www.openarchives.org/OAI/openarchivesprotocol.html>

23 <http://www.openarchives.org/index.html>

vester basiert auf XML und verwandten Standards, wie beispielsweise SAX²⁴ (Simple API for XML)

Ähnlich wie Apache's Cocoon²⁵, ein Framework für Web Publishing zur Verarbeitung von XML, basiert das Harvesting-Konzept von SozioNet auf der Nutzung von Pipelines. Während bei Cocoon ein XML-Stream eingegeben und ein beliebiges Datenformat ausgegeben wird, funktioniert das SozioNet-Konzept genau andersherum: Ein beliebiges Format kann eingesammelt werden, durchläuft verschiedene Stationen, z.B. Normalizer, Summarizer, Serializer und wird als wohl geformtes XML ausgegeben.

Institute, die die Metadaten erstellt haben, geben die Adressen der zu durchsuchenden Seiten an, und die Harvesting-Komponente von SozioNet durchläuft diese dann in regelmäßigen Abständen.

Alle eingesammelten Metadaten werden in einer zentralen Datenbank gespeichert. Der Harvester übernimmt die Volltextindexierung und muss auf Updates reagieren. URLs können auch direkt in die Datenbank eingegeben werden, so dass der Harvester nicht vollständig die Webadressen indiziert, sondern nur solche, bei denen bekanntermaßen für SozioNet relevante Ressourcen vorhanden sind.

Da alle Komponenten in SozioNet in XML beschrieben sind, werden die Metadatenätze in einer XML-Datenbank vorgehalten.

Die folgende Abbildung gibt einen schematischen Überblick über die Architektur von SozioNet:

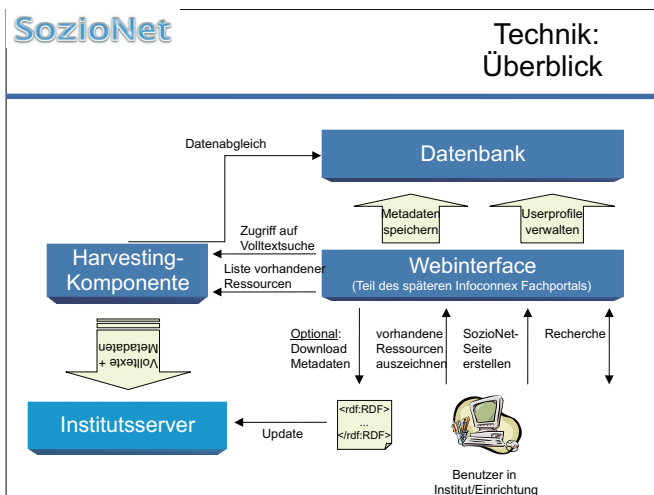


Abb.1: Architektur von SozioNet

24 <http://www.saxproject.org/>

25 <http://cocoon.apache.org/>

6 Aufgaben der Projektpartner

Derzeit beteiligen sich elf Einrichtungen an SozioNet. Sie werden ihre Ressourcen so aufbereiten, dass diese über SozioNet recherchierbar sein werden. Außerdem werden sie sich an der Weiterentwicklung des Metadatenschemas beteiligen und die bereitgestellten Werkzeuge evaluieren.

Bei den Projektpartnern handelt sich sowohl um universitäre Institute als auch um außeruniversitäre Forschungseinrichtungen. Einige Einrichtungen, wie beispielsweise das Deutsche Jugendinstitut in München oder das Wissenschaftszentrum Berlin für Sozialforschung, verwalten ihre Materialien in eigenen Content-Management-Systemen, während andere erst damit beginnen, ihre Dokumente im Netz zu veröffentlichen.

Dementsprechend ergeben sich für die Projektpartner unterschiedliche Aufgaben. Einrichtungen mit eigenen Systemen müssen entsprechende Schnittstellen zur Verfügung stellen, damit SozioNet auf die vorhandenen (Meta-)Daten zugreifen kann. Es bietet sich an, hier das Open Archives Initiative (OAI) Protocol zu verwenden. Über ein standardisiertes Protokoll (OAI-MHP) können verteilte Metadaten abgefragt und eingesammelt werden. Außerdem müssen Schnittstellen zu den Erschließungswerkzeugen (Klassifikation und Thesaurus) des IZ Sozialwissenschaften entwickelt werden. Der SozioNet-Harvester sammelt die Metadaten dann ein.

Anders sieht die Aufgabenstellung aus, wenn Ressourcen nicht gezielt verwaltet und noch nicht mit Metadaten ausgezeichnet worden sind. In diesem Fall müssen die entsprechenden Dokumente mit Metadaten gemäß der SozioNet Ontologie versehen werden.

Da in der Regel bei Sozialwissenschaftlerinnen und Sozialwissenschaftlern Kenntnisse über Metadatenformate nicht vorausgesetzt werden können, haben wir ein Formular als Eingabehilfe (MetaWizard) für die Auszeichnung von Online-Ressourcen mit Metadaten entwickelt. In diesem webbasierten Formular spiegelt sich das zugrunde liegende SozioNet Metadatenschema wider. Die Projektbeteiligten tragen in den MetaWizard die entsprechenden Daten ein und dann wird automatisch ein XML-Metadatensatz erzeugt.²⁶

Die Erstellung des Metadatensatzes beginnt mit der Angabe der vollständigen URL der zu erschließenden Ressource. Bereits vorhandene Metadaten wie beispielsweise HTML-Metatags werden automatisch extrahiert und in das Formular übernommen. Diese müssen dann nicht mehr zusätzlich eingegeben werden, sondern möglicherweise nur noch angepasst werden.

Neben der formalen Erschließung sollen vor allem auch inhaltliche Merkmale beschrieben werden. Das geschieht zum einen durch die Vergabe von frei wählenden Schlagwörtern und zum anderen über die entsprechenden Einträge

²⁶ Es besteht auch die Möglichkeit, eigenständig einen dem Metadatenschema von SozioNet entsprechenden Metadatensatz zu erstellen.

aus dem kontrollierten Vokabular des Thesaurus und der Klassifikation Sozialwissenschaften. Darüber hinaus besteht die Möglichkeit, durch ein kurzes Abstract die wesentlichen Inhalte der Ressourcen zu beschreiben.

MetaWizard

Seite 2: Allgemeine Metadaten

Titel <small>Der Titel des Dokuments (max. 75 Zeichen)</small>	<input type="text"/>
Abstract <small>Kurze Beschreibung des Inhalts des Dokuments</small>	<input type="text" value="SozioNet ist Teil eines im Rahmen von Infoconnex entstehenden Fachportals Sozialwissenschaften, welches die vorhandenen und zukünftigen Informationsdienstleistungen des Faches bündelt und ergänzt zu den bereits bestehenden Informationsangeboten"/>
Autor(en) <small>Format: Nachname, Vorname(s). Zur Eingabe mehrerer Autoren bitte auf das ***-Symbol klicken.</small>	<input type="text"/>
Herausgeber <small>Die für die Veröffentlichung des Dokuments verantwortliche Person oder Institution</small>	<input type="text"/>
Freie Schlagworte <small>Ein oder mehrere frei gewählte Schlagworte zur Beschreibung des Dokuments. Bitte jedes Schlagwort in einem gesonderten Feld eingeben. Um weitere Felder zu erhalten, klicken Sie bitte auf das ***-Symbol.</small>	<input type="text"/>
Entnommen aus <small>Ist das Dokument die elektronische Kopie eines Originals, tragen Sie hier bitte Angaben zum Original ein.</small>	<input type="text"/>
Erstellungsdatum <small>Erstellungsdatum des Dokuments. Bitte Datumsangaben immer in der Form TT.MM.JJJJ oder nur das Jahr (z.B. 2001) eingeben, falls der genaue Zeitpunkt nicht feststeht.</small>	<input type="text" value="06.03.03"/>
Datum der letzten Änderung <small>Wann wurde das Dokument zuletzt geändert? Zum Format der Datumsangabe siehe oben.</small>	<input type="text" value="01.01.70"/>
Sprache des Dokuments	<input type="text" value="Deutsch"/>
Medienformat <small>In welchem Format liegt das Dokument vor (z.B. PDF oder Word)?</small>	<input type="text" value="HTML"/>

Abb.2: MetaWizard: Eingabeformular zur Erzeugung von Metadaten

Nachdem alle Angaben gemacht worden sind, wird automatisch ein korrekter Metadatensatz generiert. Da jede Nutzerin und jeder Nutzer einen eigenen Account hat, können jederzeit Veränderungen an bereits eingegebenen Daten vorgenommen werden. Ebenso kann der Eingabevorgang unterbrochen und ohne inhaltliche Verluste zu einem späteren Zeitpunkt wieder aufgenommen werden.

Der Metadatensatz wird nicht in das Dokument selbst eingebettet, sondern als eigene Datei im selben Verzeichnis gespeichert.

7 Zusammenfassung und Ausblick

SozioNet ergänzt das bereits bestehende Informationsangebot innerhalb der Sozialwissenschaften. Es bietet einen schnellen und bequemen Zugriff auf kostenfreie und verteilte Ressourcen des Faches. Die Verwendung von weltweit anerkannten Standards ermöglicht eine nachhaltige Nutzung und gewährleistet, dass die über SozioNet erfassten Daten auch in Zukunft nutzbar sind.

Die Grundkonzeption ist inzwischen abgeschlossen, und die Implementierungsphase hat begonnen. Der MetaWizard kann von den beteiligten Institutionen zur Erzeugung von standardisierten Metadaten gemäß der SozioNet-Ontologie genutzt werden. Die ersten Ressourcen können über SozioNet recherchiert werden.

Die Struktur für die SozioNet-Seite ist ebenfalls festgelegt und wird den beteiligten Einrichtungen zur Verfügung gestellt, so dass man sich schnell einen Überblick über die verschiedenen Institutionen und ihre Angebote in Forschung und Lehre verschaffen kann.

Kontakt

Natascha Schumann, Wolfgang Meier, Sue Heise, Rudi Schmiede
Institut für Soziologie
Technische Universität Darmstadt
sozionet@ifs.tu-darmstadt.de

ViFaPhys – Virtuelle Fachbibliothek Physik

Ein kooperatives Dienstleistungsangebot für die Physik

Thomas Severiens, Esther Tobschall

Institute for Science Networking Oldenburg,
Technische Informationsbibliothek Universitätsbibliothek Hannover

Abstract

Die Virtuelle Fachbibliothek Physik ViFaPhys (<http://vifaphys.tib.uni-hannover.de>) wird an der TIB Hannover im Rahmen der DFG Initiative „Elektronische Publikationen im Literatur- und Informationsangebot wissenschaftlicher Bibliotheken“ aufgebaut. Im April 2002 wurde mit den Arbeiten begonnen, seit Januar 2003 ist die ViFaPhys online und soll im Dezember 2003 in den Routinebetrieb übergehen. Ziel ist der integrierte Zugang zu fachrelevanten Informationen und Dienstleistungen unabhängig von der physischen Form der Informationsquelle. Die hohe Qualität dieses Angebotes unter technisch-organisatorischen, wie inhaltlichen Gesichtspunkten soll durch die Einbeziehung von Fachspezialisten aus den Bibliotheken sowie insbesondere aus den Fachbereichen der Physik garantiert werden. Die TIB Hannover als zentrale Fachbibliothek für Physik koordiniert den Aufbau der ViFaPhys.

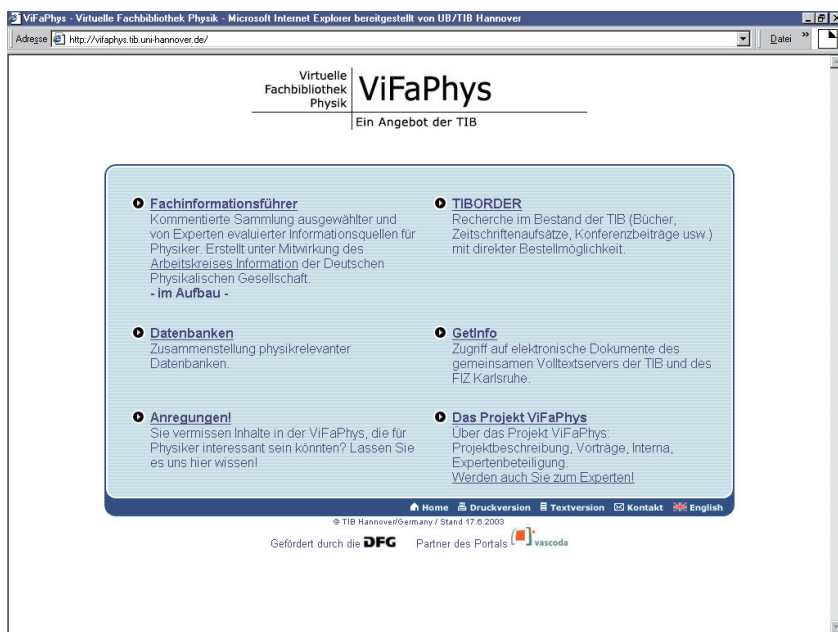
Einleitung

Aufgabe der Virtuellen Fachbibliothek Physik ViFaPhys ist es, einen integrierten Zugang zu physikrelevanten Informationen und Dienstleistungen zu bieten. Die Virtuelle Fachbibliothek Physik ergänzt insoweit das klassische Literatur- und Informationsangebot der Bibliotheken für die Physik, als dass der überwiegende Teil der erfassten und nachgewiesenen Informationsquellen über das Internet verfügbar ist. Nichtsdestoweniger ist es das Ziel, alle relevanten Angebote, unabhängig von der Art und der physischen Form der einzelnen Informationen, zu berücksichtigen.

Das wesentliche Charakteristikum der Virtuellen Fachbibliothek Physik und bisher ein Novum bei den Virtuellen Fachbibliotheken ist die direkte, organisierte Einbindung der Fachwissenschaftler, d.h. der Physikerinnen und Physiker in Hochschule und Forschung über den Arbeitskreis Information AKI der Deutschen Physikalischen Gesellschaft DPG.

Die Struktur der Virtuellen Fachbibliothek Physik orientiert sich an bereits bestehenden Angeboten, insbesondere der Virtuellen Fachbibliothek Technik¹, sie ist daher ebenfalls modular aufgebaut: Kernmodule der Virtuellen Fachbibliothek Physik sind der Fachinformationsführer und das Modul Datenbanken. Weitere Dienste vervollständigen das Angebot.

Die hohe Qualität der Inhalte und der technischen Implementierung wird durch die Bündelung der Kompetenzen aus verschiedenen Institutionen und Arbeitsschwerpunkten sichergestellt.



Die Einstiegsseite der ViFaPhys gibt den modularen Aufbau des Dienstes wieder.

Aufgrund ihrer Funktion als Deutsche Zentrale Fachbibliothek für Physik koordiniert die Technische Informationsbibliothek Hannover den Aufbau der Virtuellen Fachbibliothek Physik. Sie stützt sich hierbei auf Erfahrungen, die bereits beim Aufbau der Virtuellen Fachbibliothek Technik gesammelt werden konnten.

¹ Mit den Arbeiten zum Aufbau der Virtuellen Fachbibliothek Technik wurde im April 1999 begonnen. Seit April 2000 ist sie im Netz und unter <http://vifatec.tib.uni-hannover.de> zu erreichen.

Der Arbeitskreis Information (AKI)² der Deutschen Physikalischen Gesellschaft organisiert – auch technisch – die fachwissenschaftliche Begutachtung der Inhalte des Fachinformationsführers der ViFaPhys. Hierzu werden in Kooperation mit der Konferenz der Fachbereiche Physik (KFP) Wissenschaftlerinnen und Wissenschaftler aus den Fachbereichen und Instituten als Gutachter gewonnen, die dann in einem weitestgehend automatisierten Prozess, die Quellen beschreiben und bewerten sowie fachlich einordnen.

Durch Kooperation mit dem Fachinformationszentrum Technik³ wird die Einbindung der Datenbank INSPEC in das Angebot der ViFaPhys ermöglicht.

Das Institute for Science Networking (ISN)⁴ an der Universität Oldenburg programmiert und implementiert die Technik hinter der Nutzeroberfläche der ViFaPhys. Dabei kann auf Erfahrungen beim Aufbau zahlreicher Physik-Portale zurückgegriffen werden, wie dem weltweiten verteilten Portal PhysNet⁵ und der Fachwelt-Physik⁶ des AKI.

Die ViFaPhys implementiert einen Workflow, der sowohl Fachwissenschaftler als auch Bibliothekare einbindet. Diese Kooperation soll langfristig die inhaltliche Qualität der ViFaPhys als Dienstleistung für die Wissenschaft sichern.

Allgemeine Anforderungen an Portale und ihre Berücksichtigung in der ViFaPhys

Der wissenschaftliche Beirat der IuK-Physik hat in 2001 Kriterien zusammengestellt, die ein Web-Portal erfüllen soll, um optimal den Bedürfnissen der Nutzer zu genügen [1]. Diese Kriterien wurden in der ViFaPhys realisiert:

1. Ein Portal soll die Nutzer zu den gesuchten Informationen leiten. Die Auswahl der Quellen muss gezielt auf den angesprochenen Nutzerkreis zugeschnitten sein. Das Portal soll keine irrelevanten Quellen nennen (precision), aber es müssen alle relevanten Quellen verzeichnet sein (coverage). Die ViFaPhys stellt durch die Begutachtung jedes Links sicher, dass nur fachrelevante Quellen enthalten sind. Ein Upload-Interface zum Vorschlagen neuer Dienste und Quellen sichert Aktualität und Coverage. Veralterte Links werden in der ViFaPhys durch regelmäßige automatisierte Linkchecks und eine wiederholte Begutachtung der verlinkten Quellen aufgedeckt und aus dem Portal entfernt.

2 <http://www.aki-dpg.de> Arbeitskreis Information (AKI) der Deutschen Physikalischen Gesellschaft

3 <http://www.fiz-technik.de> Fachinformationszentrum Technik

4 <http://www.isn-oldenburg.de> Institute for Science Networking Oldenburg GmbH

5 <http://www.physnet.net> PhysNet

6 <http://www.fachwelt-physik.de> Fachwelt Physik

2. Die Inhalte eines Portals müssen nachvollziehbar sein. Für die Auswahl und die Evaluierung der Quellen wurden Kriterien definiert, diese sind auf dem Server der ViFaPhys dokumentiert. Grundlegend für die Definition der Auswahlkriterien war die Festlegung der Zielgruppe der ViFaPhys: Das Angebot der ViFaPhys wendet sich vornehmlich an Fachwissenschaftler und Fachwissenschaftlerinnen. Die abzudeckenden Fachgebiete und die Art der zu berücksichtigenden Informationsquellen werden durch das Sammelprofil der ViFaPhys beschrieben. Nachvollziehbarkeit bedeutet auch, dass ein Nutzer einmal gefundene Inhalte in seine Bookmarks aufnehmen kann, weil jede einzelne Seite der ViFaPhys individuell adressierbar ist.
3. Ein Portal soll Informationen liefern, sich aber selbst auf die Anforderungen, die sich aus dieser Aufgabe ergeben, beschränken. Die Benutzung der ViFaPhys erfordert keinerlei spezielle Plugins, die Benutzeroberfläche ist auf das funktionell Notwendige beschränkt und unabhängig vom verwendeten Anzeigegerät benutzbar. Insbesondere gibt es eine reine Text-Version ohne Bilder und Tabellen, die für die Ausgabe auf einer Braille-Zeile geeignet ist.
4. Eine weitere allgemeine Anforderung der IuK-Physik war, dass ein Portal verschiedene Schnittstellen bieten muss. Zwei essentielle Schnittstellen wurden bereits erwähnt: Eine Schnittstelle für den Besucher mit einem grafischen Web-Browser, eine Schnittstelle für die Braille-Zeilen. Weitere sinnvolle Schnittstellen sind eine Druckansicht jeder Web-Seite des Portals, die dann beispielsweise die URL der jeweiligen Seite enthalten sollte und die URLs der gelisteten Online-Quellen im Klartext (nicht nur als unterlegter Hyperlink) angibt. In der ViFaPhys sind zwei weitere XML-basierte Schnittstellen implementiert: Das Open Archives Initiative⁷ Protocoll for Metadata Harvesting bietet die Möglichkeit, wissenschaftliche Datenquellen über ein standardisiertes Protokoll abzufragen, um ihre Inhalte in Diensten anzubieten. Jedes Portal ist in diesem Sinne als eine Datenquelle zu verstehen, die beschreibende Metadaten über die Online-Dienste anbietet. SOAP als Protokoll für Web-Services bietet die Möglichkeit, die Funktionalitäten eines Dienstes in einem anderen Dienst zu nutzen. Insbesondere die Suchfunktion eines Portals kann in thematisch breiteren Portalen integriert werden: Die ViFaPhys ist so z.B. in vascoda⁸, dem interdisziplinären Internetportal für wissenschaftliche Information in Deutschland eingebunden.

7 <http://www.openarchives.org> Open Archives Initiative

8 <http://www.vascoda.de> VASCODA

Die Module der Virtuellen Fachbibliothek Physik

In vielen Virtuellen Fachbibliotheken ist der Fachinformationsführer ein zentrales Element des Angebotes. Für den Fachinformationsführer der ViFaPhys werden Zusammenstellungen von Informationen und Informationsquellen für Forschung und Lehre zusammengetragen und als kommentierte und strukturierte Sammlung übersichtlich angeboten. Nicht nur die fachliche Qualität der gesammelten Quellen sondern auch die Aktualität der Beschreibungen wird in kurzen regelmäßigen Abständen von Fachexpertinnen und -experten evaluiert.

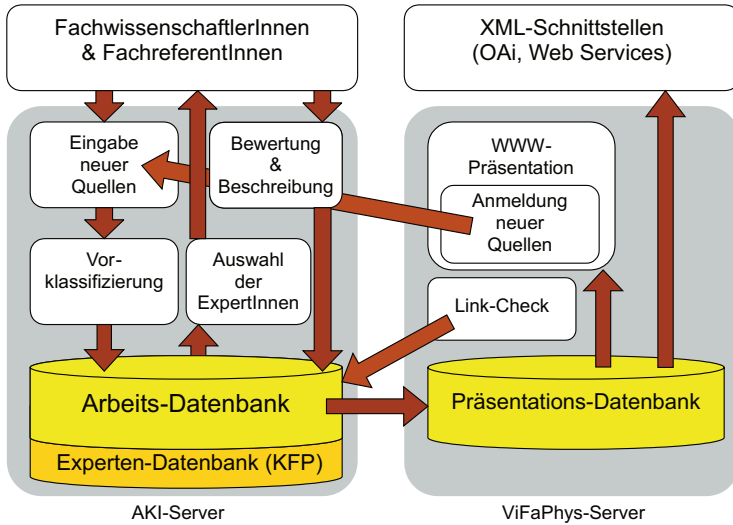
Der Fachinformationsführer enthält Linksammlungen, Volltextangebote, Faktendatenbanken, Nachschlagewerke, Softwaresammlungen, Lehrmaterialien, Links zu Institutionen und Forschungsprojekten, Veranstaltungsverzeichnisse, Mailing-Lists und Newsgroups, nicht aber Einzeldokumente. Alle Quellen sind in Fachgebiete und Kategorien des Typs der enthaltenen Information eingeteilt und lassen sich browsend oder über eine Suchmaske erschließen.

Der Fachinformationsführer wird inhaltlich durch die Einbindung von Fachwissenschaftlern und Fachwissenschaftlerinnen aufgebaut: Den Wissenschaftlern kommt die Aufgabe der Quellenauswahl, -einordnung und -beschreibung zu.

Dies und die fachliche Begleitung des Projektes insgesamt durch einen wissenschaftlichen Beirat sichert die Berücksichtigung der Bedürfnisse der Wissenschaftler und ist ein wesentliches Qualitätsmerkmal der ViFaPhys.

Die Quellen für den Fachinformationsführer werden zunächst zentral im Projekt gesammelt und als durch Experten zu evaluierende Quellen an die Arbeitsdatenbank der ViFaPhys gemeldet.

Die Abbildung veranschaulicht den Workflow des Fachinformationsführers Physik: Neue Quellen können von Fachwissenschaftlern, Fachreferenten oder jedem anderen Nutzer über ein Web-Formular gemeldet werden. Nach der Übermittlung neuer Quellen an die Arbeitsdatenbank einschließlich der Erfassung formaler Angaben, wird dann über die mitgeteilte Vorklassifizierung ein Experte zur Bewertung und Beschreibung ausgewählt. Die Arbeitsdatenbank speichert die Beschreibungen inklusive der Information über den zuständigen Gutachter. Die Beschreibungen und andere Metadaten werden wöchentlich aus der Arbeitsdatenbank in die Präsentationsdatenbank gespiegelt. Hier ist ein Link-Checking implementiert, das nicht funktionierende Links an die Arbeitsdatenbank meldet und diese ggf. sperrt. Aus den Daten der Präsentationsdatenbank werden die verschiedenen html- und xml-Schnittstellen der ViFaPhys gespeist.



Insbesondere die zeitintensive Erfassung formaler Angaben, wie z.B. zum Urheber oder Herausgeber einer Quelle kann so von bibliothekarischer Seite bereits erfolgen und muss nicht von den Experten geleistet werden. Des weiteren wird hier jede Quelle vorklassifiziert, was für den weiteren Workflow grundlegend ist.

Die Evaluation der Quellen erfolgt sowohl nach rezensierenden als auch nach webliographischen Kriterien. Die rezensierende Evaluation erfordert hierbei die Perspektive eines Fachwissenschaftlers [2].

Fachreferenten als Wissenschaftler und Bibliothekare in einer Person sind hierfür besonders geeignet, weil sie die Fachsprachen aller Beteiligten sprechen und so die Weitergabe an die fachlich bewertenden Wissenschaftler sinnvoll vorbereiten können.

Neben dem Fachinformationsführer bietet die ViFaPhys mit dem Modul Datenbanken einen direkten Zugang zu physikrelevanter Fachliteratur an:

Über eine Metasuchmaschine wird der Nutzer parallel in wichtigen Datenbanken wie z.B. ArXiv, PhysDoc oder INSPEC recherchieren können. Die Recherche und die Anzeige der Ergebnisse in einer Kurztitelliste werden frei sein, ob die Vollanzeige einzelner Treffer kostenpflichtig ist, hängt von den jeweiligen Angeboten sowie lokalen Lizenzen der Benutzer ab.

Zur Zeit findet der Nutzer im Modul Datenbanken die Liste der zukünftig parallel recherchierbaren Datenbanken.

Die ViFaPhys im Kontext anderer Portale zur Physik

Betrachtet man den Kontext der ViFaPhys, also weitere Internetangebote zur Physik, so stellt man fest, dass sich die ViFaPhys und die anderen Portale auf nationaler Ebene ideal ergänzen.

Die einzelnen Portale lassen sich durch ihre zentralen Angebote charakterisieren:

Die Fachwelt Physik⁹ richtet sich an Fachwissenschaftler und weist relevante Informationsquellen nach. Einzelne Partner der Fachwelt pflegen bestimmte Teile des Portals schwerpunktmäßig. So wird z.B. die Rubrik Fachliteratur über das Projekt ViFaPhys von der TIB betreut.

PhysNet sammelt weltweit Informationen von Physik-Fachbereichen und ihren Servern.

Das Physikportal des FIZ Karlsruhe¹⁰ stellt als Kernkomponenten die physikrelevanten Datenbanken des FIZ Karlsruhe zur Verfügung.

Das Portal der Konferenz der Fachbereiche Physik¹¹ liefert Informationen zum Physikstudium in Deutschland.

Pro-Physik¹² ermöglicht mit seiner Findemaschine eine physikspezifische Internet-Suche.

Der Fachinformationsführer der ViFaPhys ist mit seiner übersichtlichen und aktuellen Sammlung bewerteter und kommentierter Informationsquellen nutzbar als qualitätskontrollierte Linksammlung. Das Modul Datenbanken der ViFaPhys erleichtert den Zugriff auf physikrelevante Fachliteratur.

Die Welt der Physik¹³ wendet sich mit ihren Inhalten an die interessierte Öffentlichkeit.

Zusammenfassung

Für die Physik gibt es bereits zahlreiche Portale von verschiedenen Anbietern für diverse Nutzergruppen. Eine Übersicht der deutschsprachigen Portale bietet Physik.org¹⁴. Betrachtet man diese Physik-Portale einmal näher, so stellt man sehr schnell fest, dass sich nur wenige hiervon an Fachwissenschaftler wenden und einige der Portale auch nicht auf allen Browsern angezeigt werden können.

Die Virtuelle Fachbibliothek Physik (ViFaPhys) wird spätestens mit Abschluss der Implementierungsphase und dem Übergang in den Routinebetrieb Ende 2003 an den Fähigkeiten der anderen Physik-Portale gemessen werden.

9 <http://www.fachwelt-physik.de> Fachwelt Physik

10 <http://www.physik-portal.info> Physikportal des Fachinformationszentrum Karlsruhe

11 <http://www.kfp-physik.de> Portal der Konferenz der Fachbereiche Physik KFP

12 <http://www.pro-physik.de> Pro-Physik

13 <http://www.weltderphysik.de> Welt der Physik

14 <http://www.physik.org/portale.html> Auflistung von Physikportalen

Entsprechend überlegt, fachkompetent und technisch überlegen ist die ViFa-Phys zu entwickeln. Die hohe Qualität der Inhalte und der technischen Implementierung wird durch die Bündelung der Kompetenzen aus verschiedenen Institutionen und Arbeitsschwerpunkten sichergestellt. Inhaltlich wird die ViFa-Phys einen integrierten Zugang zu fachrelevanten Informationen und Dienstleistungen bieten und damit die vorhandene Portal-Landschaft sinnvoll ergänzen.

Die inhaltliche und technische Qualität des Portals soll zwar einer ansprechenden Präsentation nicht widersprechen, die Nutzbarkeit und Ergonomie stehen aber klar im Vordergrund. Die ViFaPhys soll spätestens mit Beginn des Routinebetriebes Ende 2003 zu einem wichtigen Portal für die Fachwissenschaftler und Fachwissenschaftlerinnen der Physik werden. Die aktive Einbindung der Wissenschaftler im Beirat und als Gutachter soll die fachlich hohe Qualität des Dienstes und somit eine breite Akzeptanz sicherstellen.

Literatur

- [1] <http://www.iuk-physik.de/portal/beirat.empf-dpg.html>
Empfehlungen der IuK-Physik zum Aufbau eines Physik-Portals vom 6. Juni 2001
- [2] Margo Bargheer: Qualitätskriterien und Evaluierungswege für wissenschaftliche Internetressourcen, Report zum Projekt: Datenbankbasierte Clearinghouses im Kontext digitaler Bibliotheken, Göttingen 2002

Kontakt

Thomas Severiens
Institute for Science Networking Oldenburg
severiens@isn-oldenburg.de

Esther Tobschall
Technische Informationsbibliothek Universitätsbibliothek Hannover
esther.tobschall@tib.uni-hannover.de

Weiterentwicklung von digitalen Bibliothekssystemen zu OpenArchives-Systemen

Günther Specht, Michael G. Bauer

Universität Ulm, Abt. Datenbanken und Informationssysteme,
TU München, Institut für Informatik

Zusammenfassung

Bisherige digitale Bibliothekssysteme arbeiten hauptsächlich lokal, d.h. sie halten einen lokalen Datenbestand und bieten Zugriff darauf an. Besonders für Forschergruppen ist es allerdings oft wünschenswert auf mehrere verschiedene entfernte Systeme unter einer Schnittstelle zugreifen zu können und zusätzlich mit den Dokumenten arbeiten zu können (z.B. Annotation und Verlinkung von Dokumenten). Zu diesem Zweck haben wir das System OMNIS/2 entwickelt, das diese Funktionalität bereitstellt ohne die originalen Systeme oder Dokumente zu verändern. Die Dokumente können dabei in ihren ursprünglichen Systemen verbleiben und müssen nur eindeutig auffindbar sein (sog. Searching). Neben dieser Technik hat sich in letzter Zeit, besonders durch die OpenArchives-Initiative, auch das sog. Harvesting etabliert. Dabei werden entfernte Systeme abgeerntet, die Daten lokal gespeichert und auch lokal durchsucht. Auch dieser Ansatz ließ sich in unser System integrieren, so daß dadurch den Benutzern ein deutlicher Mehrwert in der Benutzung zur Verfügung gestellt werden kann.

Abstract

Present digital library systems are usually local systems, i.e. they are based on a locally stored set of data and offer access to this data. Especially groups of researchers though ask for accessing several remote systems via one search interface and the ability to „work“ with the retrieved documents (e.g. annotation and linking of documents). To satisfy these demands we have developed the system OMNIS/2 which offers all of these features without changing the original systems or documents. The documents hereby remain in the original systems and only need to be uniquely identified for access (s.c. searching). In contrast to this technology recently another technique (s.c. harvesting) became popular mostly supported by the upcoming OpenArchives initiative. In this case the data of remote systems is harvested, the data is stored locally and also searched locally.

We were able to integrate even this approach into our system, so that users gain a real benefit by using the OMNIS/2 system.

1 Einleitung

Verschiedene Arten von Daten und Medientypen benötigen verschiedene Speicher- und Recherchesysteme. Dies führte zur Entwicklung verschiedener digitaler Bibliotheken, die mächtige, singuläre Werkzeuge darstellen, aber nicht transparent miteinander verbunden sind. So ist es z.B. nicht möglich einen Link in einem Dokument eines digitalen Bibliothekssystems zu einem anderen Dokument in der selben oder in einer anderen digitalen Bibliothek zu erzeugen oder zu verfolgen, obwohl das Quelldokument eine Referenz (etwa eine bibliographische Referenz oder ein Schlagwort) enthält. Außerdem ist es für die Benutzer nicht möglich wirklich mit den Bibliotheken zu arbeiten (d.h. Bücher zu annotieren, eigene Dokumente zu erstellen, Dokumente zu verlinken, etc.), da es sich bei den heutigen Bibliothekssystemen nur um Recherchesysteme handelt.

In diesem Papier stellen wir unser OMNIS/2-System vor¹, das es den Benutzern ermöglicht verschiedene Informationsquellen (d.h. Bibliothekssysteme) zu nutzen und sie in einer virtuellen persönlichen digitalen Bibliothek zusammenzuführen.

Das Papier ist wie folgt aufgebaut. Wir beginnen mit einer Übersicht über fehlende Fähigkeiten in aktuellen digitalen Bibliothekssystemen in Abschnitt 2. In Abschnitt 3 und 4 stellen wir die Architektur und die Implementierung von OMNIS/2 vor und erklären einige der wichtigen internen Konzepte, die Lösungen für die angesprochenen fehlenden Fähigkeiten darstellen. Abschnitt 5 beschäftigt sich mit der Erweiterung des Systems unter Einbeziehung der Open-Archives Initiative. Das Papier endet mit einer Zusammenfassung und einem Ausblick auf unsere weiteren Aktivitäten.

2 Aktuelles Szenario

Heutige digitale Bibliothekssysteme erlauben unterschiedliche Retrievalmöglichkeiten beginnend bei reinen Katalogrecherchen (OPAC-Systeme) über komplexe Volltextsuche in textbasierten Bibliotheken bis hin zu inhaltsbasierter Suche in Medienbibliotheken wie Videobibliotheken, Musikbibliotheken oder Umwelt- und Geo-Informationssystemen. Folgende Fähigkeiten werden dabei nur teilweise angeboten oder fehlen vollständig:

1 OMNIS/2 wird im Schwerpunktprogramm „Verteilte Verarbeitung und Vermittlung digitaler Dokumente“ (V3D2) von der Deutschen Forschungsgemeinschaft (DFG) gefördert.

- Die Integration von Volltextretrieval-Bibliothekssystemen und Multimedia-Datenbanksystemen zu einem fortschrittlichen, interaktiven, multimedialen Bibliothekssystem, das zusätzlich bestehende Systeme (mit oft großen Datenmengen) transparent einbinden und verknüpfen kann.
- Es besteht Bedarf an einer generischen Metadatenverwaltung, nicht nur der eigenen Dokumente, sondern auch für beliebige Dokumente zusätzlich angebundener externer Bibliothekssysteme. Dies ist notwendig zum Einsatz fortschrittlicher Filter- und Suchtechniken aber auch zur transparenten Verknüpfung und Verarbeitung von Dokumenten über Bibliothekssystemgrenzen hinweg.
- Eine interne automatische Verlinkung der Dokumente in den Bibliothekssystemen aber auch über Bibliothekssystemgrenzen hinweg. Jeder im Bibliothekssystem gefundene Artikel enthält selbst wieder Literaturreferenzen, die größtenteils im selben Bibliothekssystem gespeichert sind. Wer auch diese sehen möchte, muss heute dafür noch erneut manuell eine Suchanfrage stellen mit möglicherweise sukzessiven und damit zeitaufwendigen Einschränkungen bis zum richtigen Ergebnis. Eine automatische Verlinkung zwischen verschiedenen Dokumenten wird heutzutage nicht unterstützt.
- Zusätzlich zu automatisch generierten Links möchten Benutzer persönliche relevante Querverweise zu anderen Dokumenten im selben Bibliothekssystem als auch über Bibliothekssystemgrenzen hinweg erzeugen und verfolgen können. Insbesondere bei der Arbeit in Spezialbibliotheken wie z.B. VD17 (Verzeichnis der deutschen Drucke des 17. Jahrhunderts) [4] taucht dieser Wunsch von Forschergruppen immer wieder auf.
- In heutigen Bibliotheksrecherchesystemen fehlt größtenteils die Möglichkeit an Dokumente (seien es Text-, Bild-, Audio- oder Videodokumente) an beliebigen Stellen Annotationen zu hängen, da dies eine Schreibberechtigung für alle Nutzer in die Bibliothekssysteme voraussetzen würde. Annotationen sind aber hilfreich um sich oder einer bestimmten Benutzergruppe Inhalte zu erschließen oder zu diskutieren. Für Diskussionen sind rekursive Annotationen unterschiedlicher Sichtbarkeit erforderlich. Sollen darüber hinaus beliebige mediale Annotationen zugelassen werden, ist zusätzlich ein benutzerfreundliches Autorensystem und die Integration eines multimedialen Datenbanksystems zur Speicherung notwendig. Annotationen werden in neuentwickelten Medien- und Handbibliotheken durchaus unterstützt nicht jedoch in Katalog- und Rechercsystemen.
- Heutigen Bibliothekssystemen fehlt meist eine Personalisierungskomponente in der jeder Benutzer seine Interessen eingeben kann. Dies kann einerseits in der meist bestehenden Pull-Technologie zur zusätzlichen semantischen Informationsfilterung verwendet werden (beispielsweise dazu, Queries automatisch in einen bestimmten fachlichen Kontext zu stellen um überfachliche

Mehrdeutigkeiten per se auszuschließen) oder andererseits für eine Push-Technologie zur Benachrichtigung der Benutzer bei für ihn relevanten Neueingängen. Die Personalisierungskomponente sollte einerseits adaptiv sein, um Interessensverschiebungen der Benutzer mitzubekommen, andererseits kollaborativ um auch so etwas wie „den guten Tipp“ realisieren zu können.

3 Überblick über die Architektur

3.1 Grundsätze des Designs

Unser Ansatz stützt sich auf der Tatsache ab, dass wir bestehende digitale Bibliotheken nicht verändern können, bzw. nicht ändern möchten, da es sich bei den Systemen um etablierte Anwendungen handelt oder Änderungen einen immensen Aufwand bedeuten würden. Es ist allerdings möglich die bestehenden Systeme als große Container mit mächtigen Anfragesprachen zu betrachten. Daher bilden unsere Erweiterungen ein Metasystem für diese digitalen Bibliothekssysteme, die dadurch um neue Funktionalität erweitert werden. Unser Metasystem ermöglicht es in einer oder verschiedenen digitalen Bibliotheken zu suchen, die Dokumente automatisch zu verlinken, sie zu annotieren, sie um Multimedia-Dokumente zu erweitern und zu personalisieren. Die eigentlichen Dokumente verbleiben in den originalen Datenbanksystemen und werden nicht verändert, nicht einmal bei eingefügten Linkankern. Diese Dokumente werden in der Metadatenbank nur durch eine logische Adresse und ihre Metadaten repräsentiert. Die Verlinkung, einschließlich der Ankerpositionen wird ausschließlich im Metasystem gespeichert und wird dynamisch zur Laufzeit zu den gefunden Dokumenten hinzugefügt. Das Linkkonzept (ebenso das Dokumentenkonzept) ist mit einem Rechtekonzept versehen, das benutzerbezogene, gruppenbezogene oder allgemeinen Links berücksichtigt. Dadurch können verschiedene Sichtbarkeiten der Links (und auch der Dokumente) realisiert werden. Nach der gleichen Methode können Dokumente mit benutzerbezogenen, gruppenbezogenen oder allgemeinen Annotationen versehen werden. Die Annotationen können beliebige lange Texte, Multimedia-Annotationen oder sogar Annotationen auf Annotationen sein. Wir haben zusätzlich ein Konzept zur Personalisierung entwickelt (GRAS Algorithmus, [11]), das eine bessere Filterung und ein persönliches Ranking der Ergebnisse der unterliegenden digitalen Bibliothekssysteme bietet. Der GRAS Algorithmus verwendet im Gegensatz zu anderen Personalisierungsschemas Gauß-Kurven um Benutzer- und Objektprofile zu beschreiben, was zu einer besseren Qualität der gelieferten Antworten führt. Personalisierung kann auch als Benachrichtigungsdienst über relevante Neueingänge in der Bibliothek benutzt werden. Sowohl Push-, als auch Pull-Strategien sind hier einsetzbar. Die Annotation und die Daten, die für die Personalisierung notwendig

sind (d.h. Benutzer- und Objektprofile) werden ebenfalls in der Metadatenbank gespeichert.

Benutzer wollen oft zusätzliche persönliche Dokumente zusammen mit den originalen Dokumenten erzeugen, speichern und anfragen. Da diese persönlichen Dokumente aus Text, Tabellen, Graphiken, Audio oder Video bestehen können, muss das Metasystem selbst alle Merkmale eines interaktiven Multimedia-Datenbanksystems aufweisen, einschließlich eines Autorensystems, Volltextindizes, etc. um zusätzliche persönliche Dokumente aufnehmen zu können. Diese Dokumente werden, da sie in der Metaschicht liegen, auch als lokale Dokumente bezeichnet. Aus Konformitäts- und Konsistenzgründen werden die lokalen Dokumente aber wie Dokumente aus externen digitalen Bibliotheken behandelt (abgesehen natürlich von der Schreibberechtigung). Wenn die Dokumente annotiert werden oder mit Links angereichert werden, dann wird dies wieder außerhalb der Dokumente in der Metadatenbank gespeichert. Dieses Konzept kann als eine Erweiterung des Amsterdam Hypermedia Modells [7] gesehen werden, das selbst eine Erweiterung des Dexter Referenz Modells darstellt [6], da dabei klar die Dokumentenschicht (*within component layer*) von der Verlinkungsschicht (*storage layer*) getrennt wird. OMNIS/2 unterstützt ein sehr viel mächtigeres Linkkonzept als das augenblickliche Konzept im WWW: n:m Links, bidirektionale Links und temporale und neuentwickelte überlappende Links werden unterstützt unter Gewährleistung der referentiellen Integrität. Zusammenfassend kann das Metasystem einerseits als ein eigenständiges digitales Bibliothekssystem mit Volltext-, Multimedia- und Hypermedia-Dokumenten gesehen werden, andererseits als Metasystem für existierende digitale Bibliotheken, wobei diese um Hypermedia-Elemente, automatisch generierten und persönlichen Links, Annotationen, persönlichen Multimedia-Dokumenten, einem transparenten Zugang zu weiteren Bibliothekssystemen, einer integrierten Verwaltung von Metadaten und einer einheitlichen Benutzerschnittstelle erweitert werden.

3.2 Einbindung von externen Systemen

Im allgemeinen können zwei Möglichkeiten einer Einbindung eines bestehenden digitalen Bibliothekssystems in OMNIS/2 unterschieden werden: Eine enge Integration und eine Kopplung über XML. Bei der engen Integration werden die eingebundenen Systeme über interne Schnittstellen direkt in OMNIS/2 eingebunden. Bei der XML-Kopplung brauchen die Systeme nur um eine XML-Ausgabeschnittstelle erweitert zu werden. OMNIS/2 verwaltet dann die gemeinsame DTD (Data Type Definitions) und nutzt zusätzlich die getaggtten Informationen zur Metadatenverwaltung. Integritätsbedingungen für die Speicherung von XML-Dokumenten und DTDs in Datenbanksystemen sind daher ein wichtiger Aspekt von OMNIS/2.

4 Implementierung

4.1 Implementierungsaspekte der Architektur

Ein Kernpunkt von OMNIS/2 ist, dass weder die eingebundenen digitalen Bibliothekssysteme, noch die gespeicherten Dokumente geändert werden müssen. Es ist nicht einmal notwendig die originalen Dokumente zu ändern um Links, bzw. die Anker oder die Ansatzpunkte von Annotationen, zu kennzeichnen. Dies wird durch einen Drei-Schichten-Ansatz, ähnlich dem Amsterdam Hypermedia Referenz Modell [7], ermöglicht. Die eingebundenen digitalen Bibliothekssysteme entsprechen dem *within component layer*, der die Dokumente hält (Bild 1 zeigt die Gesamtarchitektur). Die Zusammenführung, Verlinkung und Annotationen wird in der Metaschicht durchgeführt, die aus einem Applikationsserver und einer Datenbankverbindung zur Metadatenbank besteht. Diese Schicht repräsentiert die Speicherschicht (*storage layer*) des Dexter- und Amsterdam Referenz Modells. Bei der dritten Schicht handelt es sich um die Präsentationsschicht, die sich im Webbrowser des Benutzers befindet. Damit ist die konzeptuelle Sicht auf die Architektur beschrieben (in Bild 1 dargestellt).

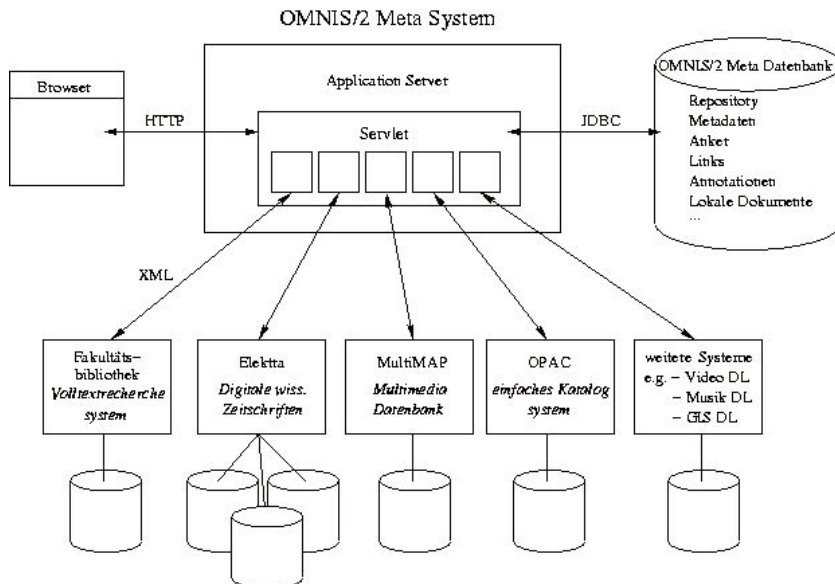


Abbildung 1: Architektur von OMNIS/2

In der mehr technischen Sichtweise der OMNIS/2-Architektur, in Abbildung 1 oben horizontal dargestellt, folgen wir bei der Realisierung der üblichen „three

tier architecture” zur Anbindung von Datenbanken ans Web. Wir verwenden dabei eine serverseitige Anbindung der OMNIS/2-Datenbank unter Verwendung von Java-Servlets und einer JDBC-Kopplung. Dabei wird ein Webserver mit Hilfe der Servlets zu einem vollständigen Applikationsserver, also dem vollständigen OMNIS/2-Kernsystem, ausgebaut. Browser und HTTP-Server kommunizieren über einen Standard-HTTP-Port, bzw. über den direkten Austausch von Java-Klassen, sofern Applets zur Anzeige nötig sind. Vom Browser kommende Benutzeranfragen werden im Servlet verarbeitet, das für den Zugriff auf die OMNIS/2-Metadatenbank eine JDBC-Kopplung zum relationalen Datenbanksystem TransBase [12] offen hält. Aus Gründen der Performanz hält das Servlet einen Pool an offenen Verbindungen bereit, aus dem nach Bedarf Verbindungen entnommen und wieder zurückgestellt werden.

Bei der Suche in einer der angeschlossenen, externen Bibliotheken muss die Suchanfrage in das entsprechende URL-Format übersetzt und versandt werden. Bei einer Linkverfolgung in eine angebundene Bibliothek muss aus den Informationen in der OMNIS/2-Metadatenbank erst die entsprechende Query im entsprechenden URL-Format des Zielsystems vom Servlet generiert werden. Im Falle der XML-Anbindung wird die XML-Antwort mittels eines XML-Parsers geparkt und um die Zusatzinformationen aus der OMNIS/2-Metadatenbank (Verlinkung, Annotationen etc.) für die Präsentation ergänzt. Dazu wird entsprechend den Ergebnisdokumenten eine zweite Anfrage an die OMNIS/2-Metadatenbank gestellt. Diese ist immer nötig. Die Ergebnisdokumente werden in OMNIS/2 als Objekte realisiert [3], ebenso die Anker und Links [2]. Alle Objekte kreieren sich selbst aus der Datenbank und können sich selbst darstellen (d.h. Dokumentobjekte kennen ihre Links und Anker). Durch diese Kapselung wird es sehr leicht das System um neue Dokumenttypen zu erweitern, da die Änderungen nur kleine Teile des Systems betreffen. Auch Autorensysteme können leicht mit einer solchen Repräsentation arbeiten. Zur Darstellung werden die Dokumente mittels XSLT nach HTML gewandelt. Falls in Zukunft Bedarf an anderen Ausgabeformaten entstehen sollte, kann dies durch ein einfaches Ersetzen der entsprechenden XSLT-Anweisungen umgesetzt werden.

In diesem Zusammenhang spielt die Architektur der Metadatenbank eine wichtige Schlüsselrolle. Dort werden Metadaten, Link-Anker, Annotationen bestehend aus verschiedenen Medientypen, aber auch lokale Dokumente, die auch aus verschiedenen Medientypen bestehen können, gespeichert. Wir erwarten, daß die Menge der gespeicherten Daten stark anwachsen wird, wenn Anker- und Linkinformation zu jedem einzelnen Objekt, das von einem Benutzer editiert wurde, gespeichert wird. Zusätzlich wird für die Personalisierung die Speicherung von Objektprofilen für alle Dokumente in der Metadatenbank benötigt. Dabei handelt es sich um stark strukturierte Daten. Die darunter liegenden Systeme verwenden in vielen Fällen aber auch semi-strukturierte Metadaten (z.B. Metadaten in XML). Die Metadatenbank übernimmt dann für die semistrukturelle

rierte Daten auch Speicherung, Suche, interne Repräsentation, Zerschlagung und Indexierung. Da sich das Metasystem auf verschiedene darunter liegende Systeme abstützt und eine eigene Metadatenbank unterhält, ist die Konsistenz zwischen den Systemen, sowohl für Metadaten, als auch für die darunter liegenden digitalen Bibliotheken ein wichtiger Aspekt.

4.2 Das Autorensystem

Um nun Benutzern von OMNIS/2 die Möglichkeit zu bieten mit den Dokumenten in gewohnter Weise arbeiten zu können, wurde ein Autorensystem entwickelt. Das Autorensystem ermöglicht es den Benutzern Dokumente mit anderen (ggf. externen Dokumenten) zu verlinken und Dokumente zu annotieren. Außerdem ermöglicht das Autorensystem das Hochladen von benutzerdefinierten Dokumenten in das System. Bei der Entwicklung wurde großer Wert auf leichte Bedienbarkeit gelegt. Nach der Untersuchung verschiedener Ansätze haben wir uns daher entschieden das Autorensystem als Java-Applet zu realisieren und damit eine Lösung auf Clientseite anzubieten. Das objektorientierte Design des Systems ermöglichte es hierbei den nötigen Datenaustausch zwischen dem Autorensystem und dem Kernsystem gering zu halten und ein zügiges Arbeiten zu ermöglichen. Durch die grafische Unterstützung ist es sehr leicht möglich Links zu setzen und Linkziele festzulegen.

Abbildung 2 zeigt einen Ausschnitt des Autorensystems. Im Text grafisch markiert ist der benutzerdefinierte Ausgangspunkt eines Links. In einem nächsten Schritt (in der Grafik nicht sichtbar) wird der Benutzer dann das Linkziel festlegen.

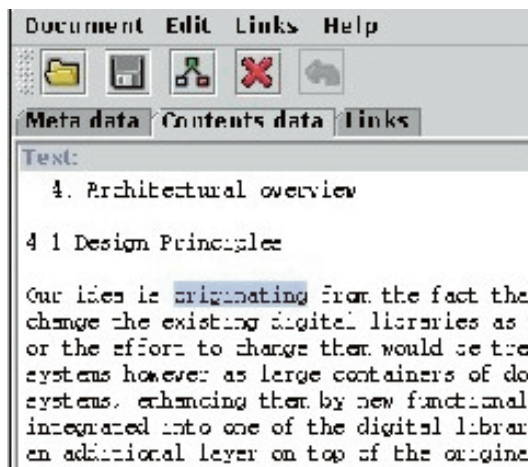


Abbildung 2: Ausschnitt des Autorensystems

5 Erweiterung zum OpenArchives Service Provider

5.1 OpenArchives

Die OpenArchives Initiative (OAI) [8, 9], hat sich zum Ziel gesetzt Interoperabilität für digitale Bibliotheken zu etablieren. Die OpenArchives Initiative entstand an der Carnegie Mellon University aus einem Vorschlag von Carl Lagoze und Herbert van de Sompel. Ziel der Initiative ist die Öffnung von digitalen Bibliothekssystemen nach außen, so daß beliebige andere Systeme auf die angeboten Daten zugreifen und nutzen können. Die ersten Vorschläge der OAI beschäftigten sich ausschließlich mit der Öffnung von bestehenden e-print Archiven (Graue Literatur, etc.). Bedingt durch den großen Erfolg der ersten Vorschläge wurde das zugrundeliegenden Material aber erweitert und umfaßt inzwischen eine breite Ausrichtung auf Content aller Art. Auffällig ist in der OAI die starke Trennung in sog. Data-Provider und Service-Provider. Als Data-Provider werden hierbei die Systeme bezeichnet, die Daten vorhalten und sich nach außen über die Möglichkeiten der OAI geöffnet haben. Im Gegensatz dazu bieten Service-Provider Dienste auf Basis dieser Daten an. Die Service-Provider ernten dazu die Metadaten bestehender Data-Provider ab (sog. harvesting) und speichern sie lokal (sog. mirroring). Die abgeernteten Daten können dann lokal durchsucht werden und Dienste darauf angeboten werden.

Hier werden Ähnlichkeiten zu OMNIS/2 deutlich. Beide Vorschläge nutzen bestehende Datenbestände um darauf Dienste bzw. einen Mehrwert für Endbenutzer anzubieten. Beide Ansätze nutzen XML als Austauschformat. Trotz der Ähnlichkeiten gibt es aber auch eine Reihe von starken Unterschieden. So ist die OAI auf Metadaten-Harvesting ausgerichtet, während in OMNIS/2 die Daten in den bestehenden Systemen verbleiben und nur bei Bedarf zur Laufzeit angefordert werden. Ein weiterer Unterschied ist die Ausrichtung von OAI auf eine Weiterentwicklung des Dienst-Protokolls aus dem NCSTRL-Projekt. Der mit Sicherheit größte Unterschied betrifft die grundsätzliche Ausrichtung beider Ansätze. Die OAI schlägt Standards vor um bestehende Archive nach außen zu öffnen, legt sich aber auf keine Implementierung fest. Hingegen handelt es sich bei OMNIS/2 um eine konkrete Implementierung, die Ähnlichkeiten zu OpenArchives kompatiblen Systemen aufweist.

5.2 Umsetzung von OAI in OMNIS/2

Aufgrund der deutlichen Ähnlichkeit in den Konzepten haben wir uns entschieden OMNIS/2 als Service-Provider für die OAI auszurichten. Um OAI-kompatible Data-Provider abzuernten, haben wir hierbei den OAI-Harvester von Jeff Young (OCLC) [1] eingesetzt. Der OAI-Harvester bietet die Möglichkeit die bei einem OAI-Data-Provider abgeernteten XML-Daten vor einer lokalen Speiche-

rung beliebigen Operationen zu unterziehen. Wir nutzen dies um die Daten für eine Ablage der relationalen Datenbank von OMNIS/2 vorzubereiten. In der relationalen Datenbank von OMNIS/2 werden die abgeernteten Daten indexiert um einen effizienten Zugriff darauf zu ermöglichen. Alle Suchanfragen auf OAI-Daten werden lokal auf diesen abgeernteten Datenbeständen durchgeführt. Ähnlich zur bisherigen Vorgehensweise wird dann auf Dokumente aus dem lokalen Datenbestand zugegriffen und diese zur Laufzeit aufbereitet (annotiert, verlinkt) und dem Benutzer präsentiert.

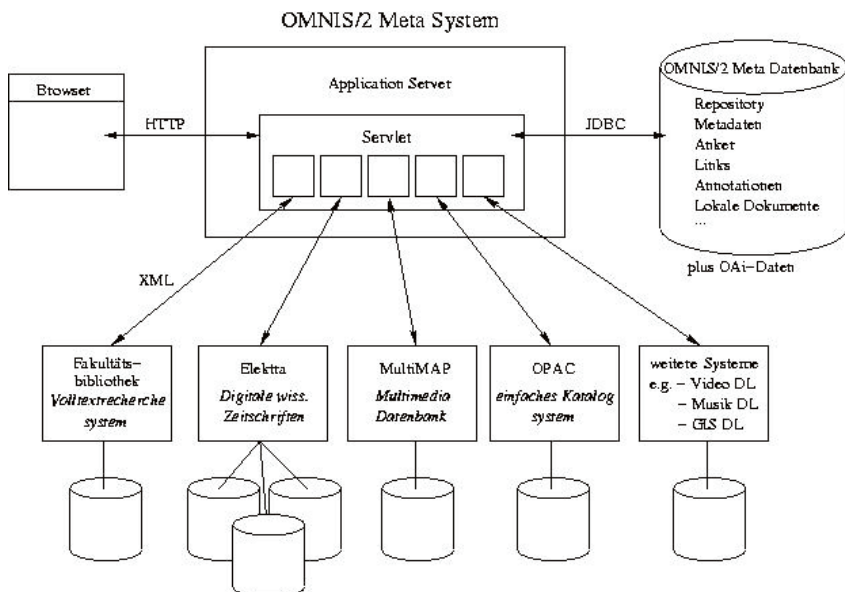


Abbildung 3: Architektur von OMNIS/2 als OpenArchives-Service-Provider

5.3 Verwandte Systeme im OpenArchives Umfeld

Speziell im Umfeld der OAI ergeben sich Ähnlichkeiten zu anderen Systemen wie z.B. dem CYCLADES Projekt. Das CYCLADES Projekt wurde etwa zwei Jahre nach dem Beginn des OMNIS/2 Projektes gestartet mit der Zielsetzung „to create a distributed and open collaborative virtual archive environment based on open archives“ [5]. Das System wurde speziell für den Informations- und Wissensaustausch in Communities (also Gruppen) entwickelt. Ein ähnlicher Ansatz ist in OMNIS/2 mit dem Rechtekonzept verwirklicht worden, bei dem

ebenfalls Dokumente und auch Links eingeschränkten Sichtbarkeiten unterliegen und Gruppensichtbarkeiten realisiert werden können.

Die starke Trennung in Daten-Anbieter und Dienste-Anbieter, die in OMNIS/2 durchgängig realisiert wurde, wird auch in CYCLADES aufgegriffen (vor dem Hintergrund der OpenArchives-Initiative) und um die Verteiltheit des Dienst-Anbieter-Bestandteils erweitert. Dies ermöglicht in CYCLADES die Aufspaltung der Funktionalität auf mehrere unterschiedliche Systeme, birgt aber natürlich höhere Anforderungen bei Ausfallsicherheit und Koordination im Allgemeinen.

Neben CYCLADES sind in der OpenArchives Initiative aber auch andere Systeme entstanden, die allerdings nicht persönliche Arbeitsumgebungen für digitale Bibliotheken als Schwerpunkt besitzen. Hier ist etwa „Callima“ zu nennen, das eine Suchschnittstelle für verschiedene OpenArchives Data Provider bietet oder auch das „iCite“-System, das ein Indexierungssystem für Literaturreferenzen von Physik-Journalen anbietet. Eine vollständige Liste der registrierten OAI-Service-Provider findet sich auf der OAI-Webseite [9].

6 Zusammenfassung

Wir haben das Konzept und die Architektur von OMNIS/2 vorgestellt. OMNIS/2 ist ein Metasystem für verschiedenste digitale Bibliotheken und ermöglicht es den Benutzern verschiedene digitale Bibliotheken auf einmal zu durchsuchen und von Links zwischen Dokumenten aus verschiedenen digitalen Bibliothekssystemen zu profitieren. Zusätzlich ist es möglich externe Dokumente zu annotieren. Beides wird ermöglicht ohne über eine Schreibberechtigung in den Bibliothekssystemen zu verfügen, die die originalen Dokumente enthalten. Erst zur Laufzeit (also wenn die Dokumente aus den darunter liegenden Bibliothekssystemen entnommen worden sind) werden die Dokumenten mit Links oder Annotationen versehen. Benutzer können außerdem durch Personalisierung ihre eigene Sicht auf die Dokumente erzeugen und so selbst mit den digitalen Bibliothekssystemen und den darin enthaltenen Dokumenten „arbeiten“, wie sie es von Dokumenten in Papierform gewohnt sind.

Wir haben ebenfalls gezeigt, wie sich die Konzepte von OMNIS/2 mit den Konzepten der OpenArchives-Initiative vergleichen lassen und sehr gut harmonisieren. Dies führte dazu, dass wir OMNIS/2 prototypisch als Serviceprovider der OpenArchives-Initiative ausgerichtet haben.

7 Literaturverzeichnis

1. ALCME, Advanced Library Collection Management Environment, OAIHarvester, <http://alcme.oclc.org/>.
2. M. G. Bauer and G. Specht. The Anchor and Linking Concept of a Meta System for Existing Digital Libraries. In *NetObjectDays 2000 - Object-Oriented Software Systems*, pages 260—265, Erfurt, Germany, Oct. 2001.
3. M. G. Bauer and G. Specht. The Object Oriented Document Model of a Meta System for Existing Digital Libraries. In *12th International Workshop on Database and Expert Systems Applications (DEXA 2001)*, 3-7 September 2001, Munich, Germany, pages 933—936. IEEE Computer Society, 2001.
4. M. Dörr, H. Haddouti, and S. Wiesener. The German National Bibliography 1601-1700: Digital Images in a Cooperative Cataloging Project. In *4th International Forum on Research and Technology Advances in Digital Libraries (ADL '97)*, May 7-9, 1997, Washington, DC, pages 50—55. IEEE Computer Society, 1997.
5. T. Gross. CYCLADES: A Distributed System for Virtual Community Support Based on Open Archives. *11th Euromicro Conference on Parallel, Distributed and Network-Based Processing (Euro PDP 2003)*, 5-7 February 2003, Genova, Italy, pages 484—491. IEEE Computer Society, 2003.
6. F. G. Halasz and M. D. Schwartz. The Dexter Hypertext Reference Model. *Communications of the ACM*, 37(2):30—39, February 1994.
7. L. Hardman, D. C. A. Bulterman, and G. van Rossum. The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model. *Communications of the ACM*, 37(2):50—62, February 1994.
8. C. Lagoze and H. V. de Sompel. The Open Archives Initiative: Building a Low-Barrier Interoperability Framework. *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2001, Roanoke, Virginia, USA, June 24-28, 2001, Proceedings*, pages 54—62. ACM, 2001.
9. The Open Archives Initiative, <http://www.openarchives.org/>.
10. G. Specht and M. G. Bauer. Omnis/2: A multimedia meta system for existing digital libraries. In *Research and Advanced Technology for Digital Libraries, 4th European Conference, ECDL 2000, Lisbon, Portugal, September 18-20, 2000, Proceedings*, volume 1923 of *Lecture Notes in Computer Science*, pages 180—189. Springer, 2000.
11. G. Specht and T. Kahabka. Information Filtering and Personalization in Databases Using Gaussian Curves. *2000 International Database Engineering and Applications Symposium, IDEAS 2000, September 18-20, 2000, Yokohama, Japan, Proceedings*, pages 16—24. IEEE Computer Society, 2000.
12. The Relational Database System Transbase, <http://transaction.de>.

Kontakt

Günther Specht

Abt. Datenbanken und Informationssysteme, Universität Ulm

Oberer Eselsberg, D-89069 Ulm, Germany

specht@informatik.uni-ulm.de

Michael G. Bauer

Institut für Informatik, TU München

Boltzmannstr. 3, D-85747 Garching bei München, Germany

bauermi@in.tum.de

Zeitliche Verwaltung XML-basierter Metadaten in digitalen Bibliotheken

Markus Kalb, Günther Specht

Universität Ulm, Abt. Datenbanken und Informationssysteme

Zusammenfassung

In digitalen Bibliotheken hat sich die Verwendung von XML als Austauschformat für Metadaten etabliert. Zunehmend kommt die Anforderung hinzu, auch komplexe zeitliche Daten und Anfragen innerhalb von XML effizient zu unterstützen. Allerdings sind die heutigen Möglichkeiten der XML-Technologien für eine zeitliche Verwaltung von Daten nicht ausreichend. Das im folgenden vorgestellte Modell T-XPath erweitert das Datenmodell und die Anfragesprache XPath um eine flexible und effiziente zeitliche Verwaltung, die auch unscharfe und ungenaue zeitliche Information berücksichtigt. Zusätzlich werden verschiedene Implementierungsvarianten von T-XPath vorgestellt und ihre jeweiligen Vor- und Nachteile hinsichtlich Performanz, Effizienz und Realisierbarkeit diskutiert. Damit wird eine korrekte zeitliche Annotation von Metadaten sowohl innerhalb digitaler Bibliotheken als auch in ihren Austauschformaten erreicht.

Abstract (Extended)

The use of XML as an interchange format for meta data in digital libraries is commonly established. The requirement to manage complex temporal data and queries effectively inside XML is increasingly important. Temporal meta data arises especially from historical data or series of measurements in digital libraries and E-Journals as well as in their user management. The possibilities of XML especially of XPath and XSchema, are insufficient for a robust and flexible representation and handling of dates or rather nonexistent for a temporal management of information. In the following we present the temporal data model T-XPath that extends the data model and query language XPath by a flexible and robust temporal management. The concept is based on abstract temporal data types (ADT), which encapsulate the entire history of a value. Additionally supported are indeterminacy and fuzzy dates. The query language is downward compatible to XPath and supports temporal queries by new temporal operations and functions. Thus a correct temporal annotation of meta data is guaranteed both inside digital libraries and in its interchange formats. Consequently, it also sa-

ves the quality and history of data. In addition to the fundamentals of T-XPath, different variants for its implementation and its respective advantages and disadvantages are discussed. Thereby, the issues performance, efficiency and short-term realisation are also captured.

1 Einleitung

XML hat sich als Austauschformat für Metadaten in digitalen Bibliotheken etabliert. Zunehmend kommt die Anforderung hinzu, auch komplexe zeitliche Daten und Anfragen innerhalb von XML effizient zu unterstützen. Zeitliche Metadaten fallen insbesondere an bei zeitbasierten (z.B. historischen) Daten oder Messreihen in digitalen Bibliotheken und E-Journals sowie bei der Benutzerverwaltung in digitalen Bibliotheken. Für XML-Dokumente wurde vom W3C die Anfragesprachen XPath und (darauf aufbauend) XQuery vorgeschlagen [W3C02]. Die bisherigen Möglichkeiten von XML, insbesondere XPath und XSchema, sind für eine robuste und flexible Behandlung von Zeitangaben nicht ausreichend bzw. für eine zeitliche Verwaltung von Information nicht vorhanden. Das im folgenden vorgestellte zeitliche Modell T-XPath erweitert das Datenmodell und die Anfragesprache XPath um eine flexible und effiziente Verwaltung und Abfragemöglichkeit von zeitlichen Informationen.

Das Papier ist wie folgt gegliedert: In Abschnitt 2 werden die bisherigen Möglichkeiten von XML zur Unterstützung von zeitlichen Informationen dargestellt. In Abschnitt 3 wird nach Betrachtung der grundlegenden Aspekte einer zeitlichen Verwaltung, das temporale Datenmodell und die Anfragesprache, T-XPath, vorgestellt. In Abschnitt 4 werden verschiedene Implementierungssätze für T-XPath vorgestellt und diskutiert.

2 Unterstützung von zeitlichen Informationen in XML

Bei der Betrachtung der bisherigen Möglichkeiten zur Unterstützung von zeitlichen Informationen in XML liegt der Fokus auf den Zusatztechnologien zu XML, XSchema und XPath. XPath ist eine Anfragesprache zur Navigation innerhalb der hierarchischen Struktur eines XML-Dokumentes, zur Selektion von Teilen des Dokumentes sowie zur Manipulation der selektierten Daten [W3C02]. Im Folgenden wird die derzeit noch in Arbeit befindliche Version 2.0 zugrunde gelegt. In XPath werden Zeitangaben durch die Verwendung der einfachen zeitlichen Datentypen aus XSchema definiert: *duration*, *dateTime*, *time*, *date*, *gYearMonth*, *gYear*, *gMonthDay*, *gMonth*, *gDay*. Diese bieten lediglich eingeschränkte Möglichkeiten zur Behandlung von zeitlichen Angaben. So sind die einzelnen Datentypen entsprechend dem gregorianischen Kalender defi-

niert. Jeder Datentyp entspricht einer anderen Granularität des Kalenders. Das hat vielfältige Nachteile: Vergleiche zwischen den unterschiedlichen Datentypen werden dadurch erschwert, Änderungen des Datenformates oder des Kalenders gestalten sich aufwendig, beliebige zeitliche Perioden, wie beispielsweise „03.10.01 - 06.10.01“, müssen explizit modelliert werden. Des weiteren werden Aggregationen (Coalescing) von zeitlichen Perioden nicht direkt unterstützt. Ebenfalls problematisch erweist sich die Abhängigkeit der Operationen von dem Datumsformat der Zeitangaben (z.B. `add-year- MonthDuration-to-dateTime(A,B)`). Schließlich sind ungenaue und unscharfe Zeitangaben nicht definierbar.

Die bisherigen Möglichkeiten von XPath und XSchema sind für eine robuste und flexible Behandlung von Zeitangaben nicht ausreichend bzw. für eine zeitliche Verwaltung von Information nicht vorhanden. Das im folgenden vorgestellte Modell T-XPath, erweitert die XML-Technologie um eine effiziente Verwaltung von komplexen zeitlichen Informationen durch die Einführung neuer zeitlicher Datentypen sowie ein Konzept ihrer Interaktionen mit konkreten Werten.

3 Zeitliche Verwaltung mit T-XPath

3.1 Neue zeitliche Datentypen für Zeitangaben in T-XPath

Die drei zeitlichen Datentypen, Zeitdauer (*interval*) zur Charakterisierung der Dauer eines Ereignisses, Zeitpunkt (*instant*) für ein Ereignis mit der Dauer eines Chronons¹ und Zeitperiode (*period*) zur Angabe eines Zeitbereiches während dem ein Ereignis stattgefunden hat, bilden die Grundlage von T-XPath.

Alle zeitliche Datentypen legen Chronons zugrunde. Sie sind somit unabhängig von einem spezifischen Kalender und die Umrechnung in konkrete Kalendarien erfolgt über Abbildungsfunktionen. Zusätzlich unterstützen alle Datentypen verschiedene Formen von Unschärfen wie sie durch die Messung der Zeit oder durch unzureichende zeitliche Informationen entstehen können. Beispielsweise entspricht eine ungenaue Zeitdauer der Aussage „das Ereignis dauerte zwischen 5 und 7 Stunden“. Dies wird im Datentyp *interval* durch Minimum und Maximum von möglichen Werten angegeben. Die Einbeziehung unscharfer Zeitangaben erfordert jedoch eine Erweiterung einiger Operationen. Insbesondere boolesche Operationen müssen um einen dreiwertigen booleschen Datentyp (3-bool [CP01]) ergänzt werden, da aufgrund der Unschärfe nicht immer nach True oder False ausgewertet werden kann, sondern auch nach Vielleicht (Maybe).

¹ Ein Chronon entspricht der kleinsten nicht weiter teilbaren, diskreten Einheit einer eindimensionalen, linearen Zeitlinie.

Die drei Datentypen *interval*, *instant* und *period* sowie ihre zugehörigen Operationen (z.B. arithmetische, Vergleichs- und Cast-Operationen) ersetzen die bisherigen neun zeitlichen Datentypen von XSchema respektive XPath. Eine ausführlichere Betrachtung aller Komponenten von T-XPath findet sich in [KSS03].

3.2 Die abstrakten zeitlichen Datentypen von T-XPath

Die Grundlage der zeitlichen Verwaltung in T-XPath bilden abstrakte Datentypen (ADT). Diese Form des Attributzeitstempelverfahrens [Sk97, JE00] kapselt die gesamte Geschichte eines Wertes innerhalb eines ADTs [Er99, Gü00, CR01]. Ein ADT erweitert einen nichtzeitabhängigen Datentyp um eine zeitliche Verwaltung, wobei dessen ursprüngliche Eigenschaften bestehen bleiben und lediglich um neue zeitliche Eigenschaften ergänzt werden. Die gekoppelte Speicherung von konkreten Werten und Zeitangaben in einem Datentyp ermöglicht effiziente Algorithmen, insbesondere für Operationen, die gleichzeitig auf zeitlichen und konkreten Werten operieren (z.B. Änderungsrate eines Wertes). Diese Operationen waren mit den bisherigen Verfahren nur schwer oder nicht effizient lösbar [Er99].

Die zeitliche Verwaltung mit ADTs erweitert lediglich das Typsystem von XPath bzw. XSchema. D.h. es werden neue zeitliche Datentypen und Operationen eingeführt, die in XSchema verwendet werden können. Die Baumstruktur des zugrundeliegenden XPath- Datenmodells wird durch die zeitliche Verwaltung nicht beeinflusst. In T-XPath werden für alle bisherigen Datentypen (z.B. *string*, *integer*, etc.) drei korrespondierende abstrakte zeitliche Datentypen eingeführt, die deren Gültigkeitszeit- (*valid time*), wann war ein Wert in der realen Welt gültig, Aufzeichnungszeit- (*transaction time*), wann wurde der Wert gespeichert oder bitemporale Entwicklungsgeschichte, Gültigkeit- und Aufzeichnungszeit werden zusammen betrachtet, repräsentieren, z.B. für den Datentyp *string*: *v_string* (Gültigkeitszeit), *t_string* (Aufzeichnungszeit) und *vt_string* (bitemporale Zeit). Die jeweiligen Eigenschaften der Zeitarten, beispielsweise keine Lücken in der Aufzeichnungszeit, sind für jeden der Datentypen formal definiert und gewährleisten eine konsistente zeitliche Verwaltung. Weiterführende Eigenschaften der zeitlichen Verwaltung werden bei der Modellierung der Anwendung über zusätzliche zeitliche Eigenschaftsattribute des ADT explizit festgelegt und bei der Instanziierung von dem ADT überprüft, z.B. ob Überlappungen der zeitlichen Angaben zulässig oder nicht gewünscht sind. Auf einem ADT sind alle Operationen seines ursprünglichen Datentyps sowie alle Operationen der zeitlichen Datentypen definiert. Der Unterschied zu den originalen Operationen liegt lediglich in dem neuen Ergebnistyp. Das Ergebnis kann durch die Mengenwertigkeit der ADTs mehrere Elemente enthalten, die zusammen wiederum einen instantiierten ADT bilden. Hierfür stehen ebenfalls Operatio-

nen zur Verfügung (z.B. *count()*, *last()*, *next()*) die den Umgang mit der Mengenwertigkeit erleichtern.

3.3 Repräsentation der abstrakten Datentypen in XML und XSchema

Für T-XPath sind die abstrakten zeitlichen Datentypen in XSchema integriert und können durch den Anwender analog zu den einfachen Datentypen bei einer Modellierung verwendet werden. Im Unterschied zu diesen ist ihre Struktur komplexer. In Abbildung 1 ist beispielhaft die Anwendung eines ADTs, d.h. eines Strings mit Gültigkeitszeit (*v_string*), aufgezeigt. In dem Beispiel unterliegt die Auflage eines Buches einer zeitlichen Entwicklung.

<pre> ... <!-- Anwendungsbeispiel eines v_string --> <xs:complexType name="Buch"> <xs:sequence maxOccurs="unbounded"> <xs:element name="Titel" type="string"/> <xs:element name="Auflagen" type="v_string" value="Auflage" time="valid_time" format="MM.YYYY"> </xs:sequence> </xs:complexType> ... </pre>	<pre> ... <Buch> <Titel> Physik I </Titel> <Auflagen> <Auflage>1. Auflage</Auflage> <valid_time>02.1970-09.1975</valid_time> <Auflage>2. Auflage</Auflage> <valid_time>10.1975-05.1976</valid_time> <Auflage>3. Auflage</Auflage> <valid_time>06.1976-uc</valid_time> </Auflagen> </Buch>... </pre>
--	--

Abbildung 1: Beispiel für die Repräsentation des abstrakten Datentyp *v_string* in XSchema (links) und XML (rechts)

Auf der linken Seite in der oberen Abbildung ist die Struktur des XML-Dokumentes definiert. Dabei repräsentiert z.B. der Eintrag `<xs:element „Auflagen“ type=„v_string“...>` die Auflage eines Buches und somit den Wert der zeitlich verwaltet wird. Die Attribute `„valid_time“` und `value=„Auflage“` bestimmen, dass die Bezeichnung einer Buchauflage im XML Dokument durch das Tag `<Auflage>` und die Gültigkeitszeit des Verwendungszwecks durch das Tag `<valid_time>` gekennzeichnet ist. Das Datumsformat für Zeitangaben ist hier durch den Eintrag `format=„MM.YYYY“` definiert. Die Repräsentation des Beispiels in einem XML Dokument ist auf der rechten Seite der Abbildung dargestellt. Alle im Laufe der Geschichte aufgetretenen Auflagen eines Buches und deren jeweilige Gültigkeitszeiten sind vollständig in dem XML-Dokument enthalten.

Damit sind auch Drucke und Auflagen historischer Bücher modellierbar zu denen es keine exakte Auflagenzählung (z.B. Abdruck in „Augsburg“) und kein exaktes Auflagedatum gibt (z.B. nur bekannt, dass es im 16 Jh. war).

Die XML Repräsentation der Daten ist abwärtskompatibel zum XPath-Datenmodell. XPath könnte die zeitlichen Daten repräsentieren, allerdings mit der Einschränkung, dass keine zeitlichen Anfragen und Konsistenzbedingungen möglich bzw. überprüfbar wären.

3.4 Die Anfragesprache von T-XPath

Die Anfragesprache von T-XPath ist für nicht-zeitliche Anfragen identisch mit der Anfragesprache X-Path. Für die Verarbeitung zeitlicher Anfragen ist sie um eine Reihe neuer zeitlicher Operationen und Funktionen (siehe Abschnitt 3.2 und Abschnitt 3.6) erweitert. Die zeitliche Unterstützung von T-XPath konzentriert sich auf die Prädikate, die mittels zeitlicher Ausdrücke eine Knotenmenge weiter verfeinern. Anhand einiger Beispiele wird im folgenden die Sprache von T-XPath näher vorgestellt

Die nachfolgenden Anfragen beziehen sich auf das Beispiel von Abschnitt 3.3. Die folgende Anfrage liefert alle Bücher die im Jahre 1978 eine Auflage hatten und diese exakt bekannt war.

//Buch[Auflagen valid² ‘1978’= true]

Für den unsicheren Fall, z.B. bei historischen Büchern bei denen das Auflagedatum nur ungenau bekannt ist, müsste die Anfrage folgendermaßen umformuliert werden:

//Buch[Auflagen valid ‘1750’= maybe]

Im Ergebnis sind ausschließlich die Bücher enthalten, bei denen aufgrund unscharfer Zeitangaben für das Jahr 1750 zumindest die Möglichkeit bestand, dass sie zu dieser Zeit eine entsprechende Auflage besaßen.

Die bisher vorgestellten Anfragen wurden ausschließlich auf den zeitlichen Werten durchgeführt. Eine Anfrage, die zusätzlich konkrete Werte berücksichtigt, sieht wie folgt aus.

//Buch[Auflagen valid ‘1978’ and Auflagen = ‘3.Auflage’]

In der Anfrage wurde für das zeitliche Prädikat kein Wert angegeben. In diesem Fall wird *true* und *maybe* angenommen. Das Ergebnis bilden diejenigen Bücher,

² Die Operation valid() setzt sich aus mehreren Vergleichsoperationen zusammen und vergleicht Zeitangaben ob diese mindestens ein gemeinsames Chronon besitzen

die 1978 bereits die 3. Auflage hatten. Durch die Auswertung nach *maybe* sind auch jene Bücher enthalten von denen es nicht genau bekannt ist.

4 Implementationsaspekte von T-XPath

Für die Realisierung von T-XPath ergeben sich zwei verschiedene Varianten der Architektur und Implementierung. In der ersten Variante erfolgt eine vollständige Integration in die XML-Speicher- und Auswertssysteme bzw. Programmbibliotheken. In der zweiten Variante erfolgen die temporalen Erweiterungen nicht in den bestehenden Systemen selbst, sondern werden über eine zusätzliche Schicht bereitgestellt, welche die temporale Funktionalität auf bestehende XML-Standards abbildet. In Anlehnung an die Entwurfsprinzipien für temporale Datenbanken [Je00] wird diese Variante auch als „On Top“ Architektur bezeichnet. Beide vorgestellte Varianten haben ihre spezifischen Vor- und Nachteile die im folgenden näher betrachtet werden.

4.1 Integrierte Architektur

Bei der integrierten Architektur erfolgt eine vollständige Integration der temporalen Funktionalität in die XML-Speicher- und Auswertungssysteme. Auf Grund der unterschiedlichen Komponenten von T-XPath sind die Implementierungen der verschiedenen XML-Standards (XPath, XQuery, XSchema, XSLT etc.) von der Erweiterung betroffen. Die neuen Datentypen für Zeitangaben (z.B. *period*) und die Entwicklungsgeschichte von Werten (z.B. *t_string*) stehen dann als Basisdatentypen in der Erweiterung von XSchema zur Verfügung. Die neuen zeitlichen Ausdrücke und Operationen von T-XPath werden durch eine Erweiterung der bisherigen Implementationen der Anfragesprache XPath erreicht.

4.1.1 Vor- und Nachteile der integrierten Architektur

Der Vorteil dieser Architektur liegt in der direkten Unterstützung und Integration von zeitlichen Daten in den einzelnen XML-Komponenten. Hierdurch können konkrete Implementierungen eine effiziente Repräsentation von zeitlichen Datentypen und eine optimierte Ausführung von T-XPath Anfragen und Operationen realisieren. Nachteilig ist, dass fast alle bisherigen Komponenten und bestehenden Systeme reimplementiert werden müssen. Da solche Systeme meist nicht in der Hand des Entwicklers liegen, stellt sich eine Erweiterung dieser als sehr aufwendig oder nicht möglich heraus. Dann könnten aufgrund der Abwärtskompatibilität von T-XPath zwar T-XPath XML-Dokumente in einer Datenbank gespeichert, aber keine zeitlichen Anfragen oder Operationen an die gespeicherten Dokumente gestellt werden.

4.2 „On Top“ Architektur

Die „On-Top“ Architektur stellt die temporale Funktionalität mit Hilfe einer zusätzlichen Zwischenschicht, welche die zeitlichen Aspekte auf bestehende Techniken abbildet, bereit. Die Zwischenschicht verfügt über eine Ein- und Ausgabeschnittstelle, welche die Kommunikation zu den einzelnen Anwendungen realisiert sowie die Verbindung zu den zugrundeliegenden Techniken bildet. Die Schicht ist jedoch nicht als reiner Konvertierer zu verstehen, sondern realisiert die vollständige zeitliche Unterstützung, unter Abstützung auf ein darunter liegendes XPath-Modell.

In Abbildung 2 wird die Architekturform schematisch verdeutlicht. Die Idee der Zwischenschicht ist es auf Grundlage existierender Implementierungen von XML-Standards eine zeitliche Unterstützung für XML-Dokumenten zu realisieren. Alle zeitlichen Anfragen werden an die Zwischenschicht gestellt und durch die Schicht in X-Path konforme Ausdrücke übersetzt und an das darunter liegende XML-Dokument geleitet. Das Ergebnis wird von der Zwischenschicht entsprechend aufbereitet und als Ergebnismenge der Anfrage übergeben.

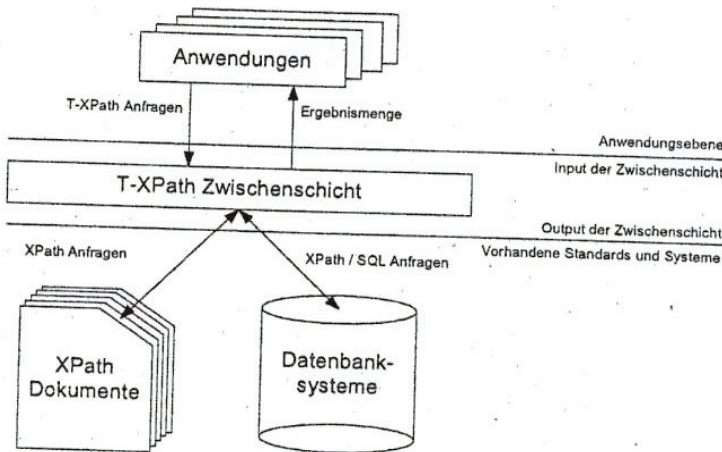


Abbildung 2: T-XPath als „On-Top“ Architektur

Die Implementierung der Zwischenschicht kann auf unterschiedliche Art und Weise erfolgen, abhängig von der jeweiligen Anwendungsumgebung:

- Die Zwischenschicht befindet sich direkt zwischen den Anwendungen und der XML-Dokumentenschicht. Somit handelt es sich um eine eigenständige

zusätzliche Schicht, welche z.B. sowohl innerhalb von digitalen Bibliotheken, als auch beim Datenaustausch zwischen diesen verwendet werden kann. Die Architekturform bietet sich insbesondere für den direkten Umgang mit XML-Dokumentensystemen bzw. XML-basierten Datenbanken an.

- Die Zwischenschicht wird in eine bereits vorhandene Middleware zwischen der Anwendung und den XML-Dokumenten integriert. Eine solche Middleware existiert z.B. bei der persistenten Speicherung von XML-Dokumenten in relationalen Datenbanksystemen und realisiert dabei eine Mappingschicht, welche ein XML-Dokument in relationale Tabellen abbildet. In diese Mappingschicht wird die zeitliche Funktionalität integriert, wobei hierbei wiederum zwei Integrationsvarianten existieren. In Variante (1) werden die T-XPath Anfragen zunächst in XPath Anfragen umgewandelt und anschließend durch die Mappingschicht z.B. in SQL-Anfragen konvertiert. In der (2) Variante erfolgt die Umwandlung unmittelbar von T-XPath Anfragen in SQL-Anfragen. Durch die Erweiterung der Middleware können dann auch Datenbanken bzw. Dokumentenformate angesprochen werden, die ursprünglich kein XML unterstützen bzw. nicht XML-konform sind.

4.2.1 Vor- und Nachteile der „On-Top“-Architektur

Aus dem „On-Top“ Ansatz ergibt sich der Vorteil, dass bereits vorhandene XML-Standardimplementierungen ohne Probleme weiterverwendet werden können und eine temporale Erweiterung ohne Modifikation der bereits bestehenden Umgebung möglich ist. Im Gegensatz zur integrierten Architektur ist mit dieser Architektur auch die persistente Speicherung von XML-Dokumenten mit zeitlichen Anfragen über diese realisierbar. Je nach Implementierungsvariante für die Zwischenschicht eignet sich die „On-Top“-Architektur sowohl für XML-basierte Austauschformate als auch für die Speicherung von zeitabhängigen XML-Daten. Allerdings führt diese Form der Architektur zu einer schlechteren Performanz gegenüber der integrierten Architektur. Dadurch, dass die verwendeten Techniken nicht modifiziert wurden, befindet sich die zeitliche Unterstützung ausschließlich in der Zwischenschicht. Die tieferliegenden Systeme besitzen keine zeitliche Unterstützung und können somit auch zeitliche Anfragen und Operationen nicht optimiert unterstützen.

5 Verwandte Arbeiten

Unsere Behandlung von Unschärfen ist eine Integration und Weiterentwicklung der Ansätze wie sie in der Literatur zu zeitlichen Primitiven zu finden sind [DS98, CP01, PT01]. Auf dem Gebiet der Repräsentation und Abfrage von zeitlichen Daten innerhalb von XML wurde bisher nur wenig Forschung betrieben. So präsentierten Grandi und Mandroli ein Modell mit expliziten Tupelzeitstem-

peln zur Beschreibung von Gültigkeitszeiten innerhalb von XML-Dokumenten [GM99]. In dem Modell wird kein Datentypkonzept verwendet, sondern die gesamte Funktionalität von zeitlichen Anfragen beruht auf parametrisierten XSL-Stylesheets. Alle zeitlichen Operationen werden als Stylesheets definiert, die das ursprüngliche XML-Dokument, unter Berücksichtigung der zeitlichen Bedingung, in ein (Ergebnis-) XML-Dokument transformieren. Es handelt sich bei diesem Vorschlag weniger um eine Integration zeitlicher Funktionalität in XML, sondern mehr um eine Modellierung einer solchen mit Hilfe von XML. In der Arbeit von Amagasa et.al. wird eine temporale Erweiterung des XPath-Modells vorgeschlagen, in der lediglich den Kanten zwischen den Elementen ein Zeitstempel zugeordnet wird und nicht den Elementen selbst [AMU00]. Wie in der Arbeit von Kalb, Schneider und Specht gezeigt, eignet sich dies jedoch besser für die zeitliche Verwaltung der Struktur eines XML-Dokuments [KSS03]. Von Dyreson werden zusätzlich zu den Kanten die Knoten mit Zeitinformation behaftet und formal in einer Erweiterung des XPath-Modells vorgestellt [Dy01]. Der Fokus liegt dabei mehr auf der Transaktionszeit und ihrer impliziten Gewinnung unter Verwendung der Änderungszeit einer XML-Datei. Ein vergleichbarer Ansatz wird von Oliboni et. al. gewählt [OQT01], wobei die Abfragefunktionalität durch eine eigene SQL-artige Anfragesprache TS-QL realisiert ist. Die vorgestellten Architekturformen wurden in ähnlicher Form in anderem Kontext von C. Jensen für die Realisierung von temporalen Datenbanken vorgestellt [Je00].

6 Zusammenfassung

Wir haben gezeigt, wie XML-basierte Metadaten in digitalen Bibliotheken zeitlich verwaltet werden können. Den Ausgangspunkt der Betrachtung bildet die Untersuchung der bisherigen Unterstützung von zeitlicher Information durch den heutigen Stand der XML-Technologie. Allerdings zeigte sich, dass sowohl die Datentypen für Zeitangaben als auch die Interaktion von Zeitangaben mit konkreten Werten den Ansprüchen einer robusten und flexiblen zeitlichen Verwaltung nicht gerecht werden konnten. Eine effizientere und flexiblere Lösung wurde durch Datenmodell und Anfragesprache T-XPath vorgestellt, dass eine kalenderunabhängige Unterstützung sowohl von exakten als auch von unscharfen zeitlichen Angaben ermöglicht. Die Anfragesprache und die ADTs von T-XPath ermöglichen eine robuste und flexible Verwaltung von zeitlicher Information, wie sie für XML-Daten bisher nicht zur Verfügung stand. Dies dient somit auch der Sicherstellung der Qualität und Historisierbarkeit von Metadaten, sowohl innerhalb digitaler Bibliotheken als auch in ihren Austauschformaten. Bei den Implementierungsvarianten von T-XPath gibt es mehrere Alternativen. Da der integrierte Ansatz Neuimplementationen aller beteiligten Systeme und

Komponenten erfordern würde und damit eine kurzfristige Bereitstellung von zeitlicher Funktionalität in XML nicht möglich wäre, bildet der „On-Top“ Ansatz die bessere Alternative. Dabei wird eine neue T-Xpath Zwischenschicht eingezo-gen, während darunterliegende Systeme und Komponenten gleich bleiben. Dieser Ansatz ermöglicht auch kurzfristig die Realisierung einer zeitlichen Unterstützung in XML.

7 Literaturverzeichnis

- [Al84] Allen, J.F. Towards a General Theory of Action and Time. Artificial Intelligence, 1984, pp. 123-154.
- [AMU00] Amagasa, T., Masatoshi, Y., Uemura, S. A Data Model for Temporal XML Documents. In Proc. Int. Conf. on Database and Expert Systems Applications - DEXA 2000, Springer Verlag LNCS 1873, London, 2000, pp. 334-344.
- [CP01] Combi, C., Pozzi, G. HMAP - A temporal data model managing intervals with different granularities and indeterminacy from natural sentences. VLDB Journal 9, 2001, pp. 294-311.
- [CR01] Chomicki, J., Revesz, P.Z. Parametric Spatiotemporal Objects. Bulletin IA*AI (Italian Association for Artificial Intelligence), Vol. 14, No.1, 2001
- [DS98] Dyreson, C.E., Snodgrass, R.T. Supporting Valid-Time Indeterminacy. ACM Trans. Database Systems 23(1), 1998, pp. 1-57.
- [Dy01] Dyreson, C.E. Observing Transaction-time Semantics with TT-XPath. Proc. 2nd Int. Conf. on Web Information Systems Engineering (WISE2001), Kyoto, Japan, 2001, pp. 193-202.
- [Er99] Erwig, M., Güting R.H., Schneider M., Vazirgiannis M. An Approach to Modeling and Querying Moving Objects in Databases. In GeoInformatica Vol.3, 1999.
- [GM99] Grandi, F., Mandreoli F. The Valid Web: it's Time to Go. TimeCenter TR-46 (www.cs.auc.dk/TimeCenter/), 1999.
- [Gü00] Güting, R.H., Böhlen, M.H., et.al. A Foundation for Representing and Querying Moving Objects. In ACM Trans. on Database Systems Vol. 25 No. 1, 2000, pp. 1-42.
- [Je00] Jensen, C.S. Temporal Database Management. PhD thesis, Univ. of Arizona, 2000.
- [KSS03] Kalb, M., Schneider K., Specht, G. T-XPath: Ein zeitliches Modell für XML-Datenbanken. In Proc. Datenbanksysteme für Business, Technologie und Web, BTW-2003, Gesellschaft für Informatik, Lecture Notes in Informatics Vol. P-26, 2003, pp. 157-166.

- [OQT01] Oliboni, B., Quintarelli, E. and Tanca, L. Temporal aspects of semi-structured data. Proc. 8th Int. Symp. on Temporal Representation and Reasoning (TIME-01), 2001.
- [PT01] Pfoser, D., Tryfona, N. Capturing Fuzziness and Uncertainty of Spatio-temporal Objects. TimeCenter Technical Report, TR-59 (www.cs.auc.dk/TimeCenter/), 2001.
- [Sk97] Skjellaug, B. Temporal Data: Time and Object Databases. Technical report, University Oslo, April 1997.
- [Sn00] Snodgrass, R. T. Developing time-oriented database applications in SQL. Morgan Kaufmann Publishers, 2000.
- [W3C02] W3C. XQuery 1.0 and XPath 2.0 Data Model (Working Draft 16.8.2002) <http://www.w3.org/TR/query-datamodel/>

Kontakt

Markus Kalb, Günther Specht
Universität Ulm
Abt. Datenbanken und Informationssysteme
{kalb,specht}@informatik.uni-ulm.de

infoconnex

Der Informationsverbund Pädagogik – Sozialwissenschaften – Psychologie

Maximilian Stempfhuber

Informationszentrum Sozialwissenschaften, Lennéstr. 30, 53113 Bonn
stempfhuber@iz-soz.de

Abstract

Der Informationsverbund Pädagogik – Sozialwissenschaften – Psychologie bietet unter seinem Markennamen infoconnex ein qualitativ hochwertiges und interdisziplinäres Informationsangebot. Basis sind die Fachdatenbanken der drei Disziplinen, die mit elektronischen Volltexten zu Zeitschriftenartikeln verknüpft werden und unter einer einheitlichen Benutzungsoberfläche recherchierbar sind. Eine besondere Herausforderung stellt die Heterogenität zwischen den Fachdatenbanken dar, die sowohl auf der Ebene der formalen wie auch der inhaltlichen Erschließung besteht. Ihre Behandlung ist die Voraussetzung für fachübergreifende Recherchen auf dem hohem Niveau, das die Nutzer von der Recherche in den einzelnen Fachdatenbanken erwarten. Infoconnex löst das Problem auf technischer Seite durch verteilte, fachspezifische Zugänge zu den Datenbanken und der Transformation der Anfragen zwischen den Datenbanken durch Transfermodule.

Was ist infoconnex?

Infoconnex¹ ist der vierte vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Informationsverbund, der als interdisziplinärer Verbund die Fächer Pädagogik, Sozialwissenschaften und Psychologie umfasst. Die Fachinformationsanbieter der drei Disziplinen und die entsprechenden Sondersammelgebietsbibliotheken, die parallel von der Deutschen Forschungsgemeinschaft (DFG) gefördert werden, arbeiten in infoconnex als gleichberechtigte Partner mit jeweils spezifischen Aufgabenschwerpunkten eng zusammen:

- Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Frankfurt

¹ <http://www.infoconnex.de>

- Informationszentrum Sozialwissenschaften der GESIS² (IZ), Bonn
- Zentrum für Psychologische Information und Dokumentation (ZPID), Trier
- SSG Bildungsforschung an der Universitätsbibliothek Erlangen-Nürnberg
- SSG Sozialwissenschaften an der Universitäts- und Stadtbibliothek Köln
- SSG Psychologie an der Saarländischen Universitäts- und Landesbibliothek, Saarbrücken

Basis der Informationsdienstleistungen von infoconnex sind die drei kostenpflichtigen Fachdatenbanken FIS-Bildung (Bildungsforschung), SOLIS (Sozialwissenschaften) und Psyndex (Psychologie). Sie werden in infoconnex mit Volltexten zu wissenschaftlichen Zeitschriftenaufsätzen verknüpft, so dass nach einer Recherche in den tief erschlossenen Datenbanken die Volltexte zu den Datenbanknachweisen bezogen werden können. Neben den kostenpflichtigen Publikationen von Verlagen wird nach Möglichkeit auch auf (kostenlose) Graue Literatur im Internet, auf Bestelldienste und auf die Elektronische Zeitschriftenbibliothek³ (EZB) an der Universitätsbibliothek Regensburg verwiesen.

Der Mehrwert von infoconnex besteht – neben der Verknüpfung von qualitativ hochwertigen, tief erschlossenen Fachdatenbanken mit Volltexten – vor allem auch in der fachübergreifenden Suche, für die im Laufe des Projekts ein ähnlich hohes Qualitätsniveau erreicht werden soll, wie dies bereits jetzt für die Recherche in den einzelnen Datenbanken der Fall ist.

Die Recherche in den Fachdatenbanken ist in infoconnex bis zur Titelliste grundsätzlich kostenlos, der Zugriff auf die vollständigen Dokumentnachweise mit Abstract und auf die Volltexte zu Verlagspublikationen ist kostenpflichtig. Für die Fachdatenbanken sind pauschale 24h- und Jahresabonnements geplant, die eine oder alle drei Fächer umfassen können. Der Zugriff auf kostenpflichtige Volltexte wird nutzungsabhängig verrechnet (pay-per-view). Spezielle Regelungen für bestehende Abonnenten der Fachdatenbanken sind zusätzlich in Vorbereitung.

Eine inhaltliche Erweiterung wird infoconnex durch parallele Projekte erhalten, zum Beispiel durch SozioNet⁴, das von der TU Darmstadt durchgeführt wird und das nach dem Vorbild von MathNet⁵ und PhysNet⁶ versucht, Metadaten für Internetressourcen zu etablieren. Daneben ist die Verzahnung mit den

2 Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (<http://www.gesis.org>)

3 <http://rzblx1.uni-regensburg.de/ezeit/fl.phtml?bibid=UBR>

4 <http://www.sozionet.de>

5 <http://www.mathnet.de>

6 <http://www.physnet.de>

Virtuellen Fachbibliotheken Sozialwissenschaften⁷ und Psychologie⁸ und dem Deutschen Bildungsserver⁹ geplant.

Aufgabenschwerpunkte im Projekt

Der Informationsverbund Pädagogik – Sozialwissenschaften – Psychologie entwickelt und betreibt den Großteil der inhaltlichen und technischen Konzepte und Komponenten selbst, greift aber auch auf externes Know-How zurück.

So wurde unter Federführung des Deutschen Instituts für Internationale Pädagogische Forschung zu Beginn des Projektes eine umfangreiche Marketingstudie in Auftrag gegeben, mit der das Kundenpotential des Volltextdienstes auf der Basis der Fachdatenbanken ermittelt werden sollte. Die Untersuchung ergab ein hohes Potential für die Dienstleistungen von infoconnex bei den anvisierten Zielgruppen in Forschung, Lehre und Praxis – gerade auch durch den hohen Bekanntheitsgrad der beteiligten Fachinformationsanbieter und ihrer Datenbanken. Auch die Wichtigkeit des direkten Bezugs von Volltexten für die Zielgruppen konnte nochmals bestätigt werden. Einen wesentlichen Faktor für die Akzeptanz des Volltextdienstes stellt dabei die Höhe des Verkaufspreises dar. Infoconnex versucht daher in seinen Verhandlungen mit Verlagen eine marktgerechte Preisgestaltung zu erreichen.

Die in einem parallelen DFG Projekt kooperierenden Sondersammelgebietsbibliotheken führen unter Federführung der Universitätsbibliothek Erlangen-Nürnberg die Verhandlungen mit den Verlagen. Ziel ist es, die für die drei Fächer relevanten deutschen und internationalen Verlage für das Projekt zu gewinnen. Vorzugsweise werden Lizenzen für den Vertrieb der elektronischen Volltexte direkt über die Server des Informationsverbundes verhandelt.

Das Informationszentrum Sozialwissenschaften entwickelt die infoconnex Portalsoftware, die von den Projektpartnern dann auf den eigenen Servern betrieben wird. Da die Datenmodelle und Datenbanktechnologien (Datenbanksysteme und Serverinfrastruktur) der drei Fachinformationsanbieter stark voneinander abweichen, galt es eine Softwarearchitektur zu realisieren, bei der ein möglichst großer Teil der entwickelten Module ohne programmiertechnische Änderungen bei allen Partnern eingesetzt werden kann. Es wurde eine container-orientierte Architektur auf der Basis der Java 2 Plattform entworfen, die unter anderem aus generischen Komponenten für Benutzungsoberflächengenerierung, Anfragerepräsentation und –verarbeitung, Authentifizierung und Abrechnung besteht. In einer späteren Ausbaustufe wird die infoconnex Software die

7 <http://www.vibsoz.de>

8 <http://fips.sulb.uni-saarland.de/port.htm>

9 <http://www.bildungsserver.de>

Weiterleitung von Anfragen zwischen den Fachdatenbanken und eine Verbundrecherche ermöglichen.

Um auch bei der Weiterleitung von Anfragen zwischen den Fachdatenbanken und bei der geplanten Verbundrecherche eine hohe Recherchequalität zu erreichen, sollen die Suchbegriffe der Benutzeranfragen mittels Cross-Konkordanzen zwischen den Thesauri der Datenbanken transformiert werden. Das Zentrum für Psychologische Information und Dokumentation erstellt hierzu schwerpunktmäßig Cross-Konkordanzen vom Psychologiethesaurus zur Pädagogik und den Sozialwissenschaften, die dann von den Partnern in der Gegenrichtung überprüft und ergänzt werden.

Infrastruktur und Architektur

Die einheitliche infoconnex Software wird bei den Partner installiert, auf die jeweilige Datenbank parametrisiert und durch Konfiguration mit einer passenden Benutzungsoberfläche versehen. Eine zusätzliche zentrale Datenbank, die alle drei Fachdatenbanken – eventuell mit einer reduzierten Erschließungstiefe – enthält, wird nicht benötigt. Zusätzlich zu den Servern bei den Projektpartnern betreibt infoconnex unter www.infoconnex.de ein Portal als zentralen Einstiegspunkt. In der ersten Projektphase verzweigt das Portal auf die fachspezifischen Unterportale, in der zweiten Projektphase wird es zusätzlich die fachübergreifende Verbundrecherche in allen angeschlossenen Datenbanken ermöglichen.

Die Fachdatenbanken werden über eine Abstraktionsschicht von den generischen Komponenten entkoppelt, spezifische Adapter übernehmen dabei die Abbildung der Nutzeranfrage auf die Schemata der einzelnen Datenbanken. Für relationale Datenbankmanagementsysteme wurde ein äußerst flexibler Adapter entwickelt, der nur durch Konfiguration an die sehr komplex strukturierten Datenbanken FIS-Bildung und SOLIS angepasst wurde. Für nicht-relationale Datenbanksysteme steht ein generischer Adapter zur Verfügung, der die Anfrage im XML-Format über das Internetprotokoll HTTP an den Server sendet. Durch die hohe Generizität der infoconnex Software ist es auch sehr einfach möglich, Anfragen zwischen den drei Installationen weiterzureichen und eine Verbundrecherche zu realisieren.

Analog zu der Anbindung unterschiedlich strukturierter Datenbanken kann infoconnex auch auf der Seite des Nutzers sehr frei an die fachspezifischen Bedürfnisse angepasst werden. Es stehen mehrere Varianten von Suchformularen zur Verfügung, die über Konfigurationsdateien für die gewünschte Suchfunktionalität parametrisiert werden können. Ein Basisdesign stellt sicher, dass der Nutzer bei allen drei Fachdatenbanken – und in einer späteren Ausbaustufe auch bei der Verbundrecherche – auf das gleiche Look-and-Feel trifft.

Die Speicherung der elektronischen Volltexte und das Abrechnungssystem wird von der Gesellschaft für Betriebswirtschaftliche Information (GBI) betrieben. Hierbei wurden Module, die für den wirtschaftswissenschaftlichen Informationsverbund EconDoc¹⁰ entwickelt wurden, nachgenutzt und an die Anforderungen von infoconnex angepasst. GBI übernimmt auch die zentrale Abrechnung aller kostenpflichtigen Dienstleistungen im Namen des Verbunds, so dass infoconnex aus Kundensicht einen One-Stop-Shop für Fachinformation und elektronische Volltexte der drei Disziplinen darstellt.

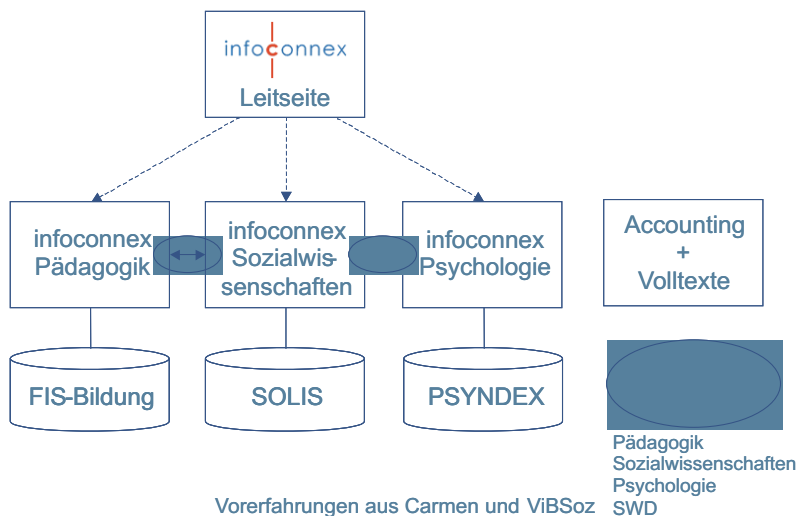


Abbildung 1: Weiterleitung von Recherchen in infoconnex

Die Abbildung 1 zeigt den schematischen Aufbau der infoconnex Infrastruktur. Beim ersten offiziellen Release, das im August 2003 stattfinden soll (infoconnex nahm bereits im April 2003 den Testbetrieb auf), wird sich der Nutzer auf der Leitseite zunächst für ein Fachgebiet entscheiden und seine Anfrage im entsprechenden Unterportal stellen. Er hat hier die Möglichkeit, die beiden anderen Fachgebiete in die Suche mit einzubeziehen. Die verwendeten Schlagwörter werden dann über Cross-Konkordanzen auf die Thesauri der anderen Fachdatenbanken abgebildet. Die Cross-Konkordanzen wurden nach dem im Projekt CARMEN definierten Verfahren erstellt; zusätzlich sollen auch statistische Verfahren zum Einsatz kommen (s. Binder et al. 2002). Die Ergebnisse aus den einzelnen Datenbanken werden zusammengeführt und dem Benutzer – nach Fächern gegliedert – angeboten.

10 <http://www.econdoc.de>

Zum zweiten Release von infoconnex, das für Ende 2003 geplant ist, soll zusätzlich eine fachübergreifende Suche vom zentralen Portal aus möglich sein. Neben den Fachthesauri wird dann auch die Schlagwortnormdatei (SWD) der Deutschen Bibliothek für die Anfrageformulierung zur Verfügung stehen. Mittels Cross-Konkordanzen werden in diesem Fall die Schlagwörter aus der Anfrage von der SWD auf die Fachthesauri abgebildet.

Benutzungsoberfläche

Zusammen mit der Bauhaus Universität Weimar und der Designfirma pool-x¹¹ entwickelte das Informationszentrum Sozialwissenschaften das Design für die Benutzungsoberfläche von infoconnex. Als Vorlage dienten unterschiedliche Informationssysteme, die im Rahmen von Forschungsprojekten wie ELVIRA (s. Stempfhuber et al. 2002) oder der Virtuellen Fachbibliothek Sozialwissenschaften (ViBSoz; s. Müller 2001) entwickelt wurden.

infoconnex

DPF UB IZ Informationszentrum Sozialwissenschaften Psychologie Information SAARLÄNDISCHE UNIVERSITÄT UND UNIVERSITÄT

Pädagogik Sozialwissenschaften Psychologie

Filter

Dokumentart

☒ Monographie ☒ deutsch

☒ Zeitschriften ☒ englisch

☒ Sammelwerk ☒ französisch

Suche

Demobetrieb

Suchformular: Bitte geben Sie hier die Begriffe ein, nach denen Sie suchen möchten. Sie können durch Leerzeichen getrennt auch mehrere Begriffe in einem Feld eingeben. Wenn alle Begriffe eines Eingabefelds in der Ergebnismenge vorkommen müssen, kreuzen Sie bitte das entsprechende Feld an.
Aus mehreren Wörtern zusammengesetzte Suchbegriffe müssen in Hochkommata gesetzt werden (z.B. 'soziale Kontrolle').

Person ☐ alle Wörter

Titel ☐ alle Wörter

Freitext ☐ alle Wörter

Schlagwort ☐ alle Wörter

Klassifikation ☐ alle Wörter

Erscheinungsjahr ☐ alle Wörter

Suche starten

Abbildung 2: Statische Suchmaske in infoconnex

¹¹ www.pool-x.de

Die Abbildung 2 zeigt einen Entwurf für die Standardsuche in infoconnex. Im oberen Bereich befinden sich Filter, mit denen die inhaltliche Suche eingeschränkt werden kann. Für die endgültige Version sind weitere Filter geplant sowie die Möglichkeit, die Filter komplett zu aktivieren und deaktivieren. Die Suche wird zunächst als statisches (Abbildung 2) und als anpassbares Formular (Abbildung 3) realisiert, die funktional ebenbürtig sind. Neben Trunkierung ist im statischen Formular sowohl die Eingabe einzelner Begriffe möglich, die dann mit dem Markierfeld „alle Wörter“ UND-verknüpft werden können, als auch die direkte Eingabe komplexer Boolescher Ausdrücke.

The screenshot displays the 'infoconnex' search interface. At the top, there are logos for 'DPF UB', 'IZ Informationszentrum Sozialwissenschaften', and 'Psychologie Information'. Below these are navigation tabs for 'Pädagogik', 'Sozialwissenschaften', and 'Psychologie'. On the left, a vertical menu lists options: 'Leistungsübersicht', 'Projektpartner', 'Speziell für Verlage', 'Impressum', 'Suche', and 'Erweiterte Suche'. The main content area is divided into two sections: 'Filter' and 'Suche'. The 'Filter' section has two columns: 'Dokumentart' with checkboxes for 'Monographie', 'Zeitschriften', and 'Sammelwerk'; and 'Sprache' with checkboxes for 'deutsch', 'englisch', and 'französisch'. The 'Suche' section is titled 'Demobetrieb' and contains a dynamic form with three rows of input fields. Each row starts with a 'und' dropdown, followed by a category dropdown (Person, Schlagwort, Klassifikation, or Jahr), and then two empty input fields separated by 'oder' dropdowns. A 'Suche starten' button is located at the bottom right of the search section.

Abbildung 3: Anpassbares Formular in infoconnex

Aus der Ergebnisliste (Abbildung 4) sind sowohl die vollständigen Dokumentnachweise als auch die Volltexte abrufbar. Die direkte Verknüpfung zu Subito und der EZB ist für registrierte Kunden ebenfalls geplant.

The screenshot shows the 'infoconnex' search interface. At the top, there are logos for DPF, UB, IZ, Informationszentrum Sozialwissenschaften, and Psychologie. Below these are tabs for 'Pädagogik', 'Sozialwissenschaften', and 'Psychologie'. A sidebar on the left contains a menu with options like 'Leistungsübersicht', 'Projektpartner', 'Speziell für Verlage', 'Impressum', 'Suche', and 'Erweiterte Suche'. The main area has a 'Filter' section with checkboxes for 'Dokumentart' (Sammelwerk, Zeitschriften, Monographie) and 'Sprache' (englisch, deutsch, französisch). Below this is an 'Anfrage' section with a search bar containing 'Schlagwort= jugendlicher und ausbildung' and a button 'Anfrage ändern'. The 'Ergebnisanzeige' section shows 'Treffer: 264' and a table of 10 results. Each result includes a checkbox, a year, a title, a database indicator (SOLIS), and a download link (subito).

	Jahr	Titel	Datenbank	Download
<input type="checkbox"/>	2002	Benachteiligte Jugendliche: Widersprüche eines sozialpolitischen Deutungsmusters : Anmerkungen aus einer europäisch-vergleichenden Perspektive	SOLIS	subito
<input type="checkbox"/>	2002	Ich muss mein Leben selber meistern! : Jugend im Stadt-Land-Vergleich	SOLIS	0,00€
<input type="checkbox"/>	2002	Öffentlich geförderte Berufsausbildung in Deutschland : Chancen der Absolventen auf dem Arbeitsmarkt	SOLIS	
<input type="checkbox"/>	2002	Jugend, Ausbildung und Beruf	SOLIS	
<input type="checkbox"/>	2002	Integration, une phase de l'insertion	SOLIS	subito
<input type="checkbox"/>	2002	Berufliche Erstausbildung in Wien : geschlechtsspezifische Segregation, Ausbildungswünsche und Berufsorientierung von Jugendlichen	SOLIS	subito
<input type="checkbox"/>	2001	Erprobung einer neuen Förderstruktur für Jugendliche mit besonderem Förderbedarf	SOLIS	subito
<input type="checkbox"/>	2001	Experiment Jung-sein : die Wertewelt österreichischer Jugendlicher	SOLIS	
<input type="checkbox"/>	2001	Jugendliche - Chance oder Problem für ländliche Räume?	SOLIS	subito
<input type="checkbox"/>	2001	Medien im Jugendalter : Rückblicke von Eltern und ihren heranwachsenden Kindern	SOLIS	subito

At the bottom of the results section, there is a button 'Nachweise anzeigen' and a page indicator '11-20'.

Abbildung 4: Titelliste mit Volltextverknüpfungen

Zusammenfassung

Der Informationsverbund Pädagogik – Sozialwissenschaften – Psychologie stellt eine wichtige Ergänzung zu den bereits bestehenden Informationsverbünden bezüglich der überregionalen Versorgung mit elektronischen Volltexten dar. Infoconnex beteiligt sich daneben am Aufbau des deutschen Wissenschaftsportals Vascoda¹², das im August seinen Betrieb aufnehmen soll. Durch die geplante enge Verzahnung mit den Virtuellen Fachbibliotheken – sowohl auf der Ebene von infoconnex als auch in Vascoda – entsteht ein zentrales, qualitativ hochwertiges Informationsangebot, das einen Großteil der Informationen der beteiligten Fächer integriert und benutzerfreundlich zugänglich macht.

12 <http://www.vacoda.de>

Literatur

- Binder, Gisbert; Marx, Jutta; Mutschke, Peter; Riege, Udo; Strötgen, Robert; Kokkelink, Stefan; Plümer, Judith (2002): Heterogenitätsbehandlung bei textueller Information verschiedener Datentypen und Inhaltserschließungsverfahren. IZ-Arbeitsbericht Nr. 24, April 2002, http://www.gesis.org/Publicationen/Berichte/IZ_Arbeitsberichte.
- Müller, Matthias N.O. (2001): Virtuelle Fachbibliothek Sozialwissenschaften (ViBSoz). In: Rützel-Banz, Margit (Hrsg.): 91. Deutscher Bibliothekartag in Bielefeld 2001 - „Bibliotheken - Portale zum globalen Wissen“. Frankfurt am Main: Klostermann. (Zeitschrift für Bibliothekswesen und Bibliographie, Sonderheft; 81). S. 37 – 43.
- Stempfhuber, Maximilian; Hellweg, Heiko; Schaefer, André (2002): ELVIRA: User Friendly Retrieval of Heterogenous Data in Market Research. In: Callaos, Nagib; Harnandez-Encinas, Luis; Yetim, Fahri (eds.): SCI 2002: The 6th World Multiconference on Systemics, Cybernetics and Informatics; July 14-18, 2002, Orlando, USA; Proceedings, Vol. I: Information Systems Development I. Orlando: TPA Publ., S. 299 – 304.

Das didaktische Metadatensystem DML

Grundlagen und praktische Anwendung

Christian Swertz

1 Einleitung

E-Learning eröffnet neue Gestaltungsmöglichkeiten für Bildungsprozesse – und erfordert zugleich die neue Gestaltung von Bildungsprozessen. Die didaktische Ontologie von Meder ist ein didaktisches Wissensorganisationskonzept, dass die mediengerechte Wissensorganisation für die Gestaltung von Bildungsprozessen didaktisch fundiert. Meder hat die didaktische Ontologie als Webdidaktik für das E-Learning spezifiziert (Meder 2003). Die Webdidaktik ermöglicht die didaktisch fundierte Entwicklung von Online-Lernplattformen und Online-Lernumgebungen. Für die technische Umsetzung wurde auf der Grundlage der Webdidaktik das didaktische Metadatensystem DML (Didactical Metadata Language) entwickelt.

Ziel der Webdidaktik ist es, durch didaktische Wissensorganisation eine Navigationsstruktur aufzubauen, die nach individuell gewähltem didaktischem Modell präsentierte Lernpfade mit Retrievalverfahren verbindet, wodurch fremd- und selbstgesteuertes Lernen ermöglicht werden. Gleichzeitig wird durch das standardisierte Metadatensystem eine effektive Inhaltsproduktion und Qualitätssicherung möglich (Swertz 2002).

Das auf der Webdidaktik basierende DML-System wurde in einer kommerziellen Plattform (SAP Learning Solution) und zwei Open-Source-Plattformen (Lerndorf, EduPlone) umgesetzt und erprobt. In diesem Aufsatz wird die Webdidaktik und das didaktische Metadatensystem vorgestellt.

2 Die Webdidaktik

Kern der Webdidaktik sind didaktische Prinzipien der Wissensorganisation. Bei Wissen handelt es sich um kulturelle Geltungsbestände, die in der Überlieferung anerkannt werden (Hönigswald 1927: 15ff.). Wissen wird im Überlieferungsprozess auf unterschiedliche Art und Weise organisiert. Dabei ist es nicht möglich, ein letztes Prinzip anzugeben, das die einzig richtige Organisation von Wissen anleiten kann, weil Prinzipien auch kulturell gebunden sind und selbst

übermittelt werden müssen. Die Webdidaktik greift daher auf kulturell bewährte didaktische Prinzipien zurück.

Ein bewährtes Prinzip ist es, die Medienwahl als didaktisches Entscheidungsfeld zu berücksichtigen (Heimann 1976). Für das E-Learning steht die Medienwahl fest: Als Unterrichtsmedium wird Computertechnik verwendet. Daher ist zunächst von den Eigenschaften der Computertechnik auszugehen. Für die Praxis des E-Learning besonders relevante Eigenschaften heutiger Computertechnik sind die Darstellung der Inhalte auf einem Bildschirm, die Universalität des digitalen Transportweges und die Möglichkeit, die Darstellung durch Algorithmen zu beeinflussen.

Die Darstellung der Inhalte auf einem Bildschirm beschränkt die Wahrnehmung der Inhalte und des Kontextes der Inhalte (Meder 1998). Am Bildschirm ist nur der Inhalt und der Kontext überschaubar, der auf einmal auf dem Bildschirm dargestellt werden kann. Dieser Effekt hat Konsequenzen für die Granularität der zu organisierenden Inhalte und die Navigation.

Die Granularität bezeichnet die Größe der kleinsten zusammenhängenden Einheit, in die Wissen für den Bildungsprozess gebracht wird. Diese Einheit wird als *Wissenseinheit* bezeichnet. Die didaktisch sinnvolle Wahl der Granularität hängt unter anderem vom verwendeten Medium ab (Meder 2003: 53). Im Falle der Darstellung auf einem Computerbildschirm ist die kleinste zusammenhängende Einheit der Inhalt, der auf einmal auf dem Bildschirm dargestellt werden kann. Für das E-Learning sind die Wissenseinheiten daher in Bildschirmgröße zu konzipieren.

Wissenseinheiten werden damit nicht im Rückgriff auf Theorien kognitiver Strukturen, d.h. als an Gedächtnisstrukturen orientierte Repräsentationen einer Wissensdomäne (Porac/Howak 2002, Walsh 1995), bestimmt. Ein solches Vorgehen ist nicht sinnvoll, da auch das Verständnis von Gedächtnisstrukturen durch kulturelle Tradierungsprozesse bestimmt wird. Bildungsprozesse sind so an kulturellen Strukturen orientiert, dass die Übermittlungsformen im Übermittlungsprozess mit erworben werden (vgl. zu den Grundlagen dieses Ansatzes: Hönigswald 1927). Daher greift die Webdidaktik auf kulturelle Tradierungsformen zurück.

Die Bewegung zwischen Wissenseinheiten ist die Abbildung von sachlogischen Bedeutungsbeziehungen in die operationale Zeit des Lernprozesses (Meder 2000a: 180). Im E-Learning wird dies als *Navigation* bezeichnet. Ein Navigationsproblem ist die Gestaltung der Schrittweite und Schrittfolge im Inhaltsraum. Da der eine Wissenseinheit umgebende Inhaltsraum am Bildschirm nicht sinnlich erfahren werden kann, muss diese Umgebung explizit dargestellt werden. Unter der Bedingung des knappen Darstellungsplatzes auf dem Bildschirm ist eine didaktisch strukturierte symbolische Repräsentation erforderlich. Diese Repräsentation erfolgt im E-Learning durch Navigationselemente.

Ausgehend von den Eigenschaften des verwendeten Mediums ist für das E-Learning also das Problem der Navigation zwischen Wissenseinheiten zu lösen. Weil mit einem digitalen Medium kaum Einschränkungen bezüglich der Darstellbarkeit bestehen, kann die Frage nach einer sinnvollen Strukturierung des Wissens aus didaktischer Sicht beantwortet werden.

In der Didaktik hat sich die Unterscheidung zwischen Inhaltswahl und Methodenwahl bewährt. Mit der Inhaltswahl wird entschieden, welche Themen zu behandeln sind. Die Entscheidungskriterien bei der Auswahl der Themen sind sachlogische und bildungstheoretische Gesichtspunkte, die zu diskutieren den Rahmen dieses Artikels überschreitet.

In der Methodenwahl wird die Anordnung des ausgewählten Wissens für den Bildungsprozess entschieden. Die Anordnung des Wissens erfolgt anhand didaktischer Modelle (z.B. entdeckend, induktiv, exemplarisch etc.). Dabei ist es erforderlich, das didaktische Modell anhand einer Analyse der didaktischen Bedingungsfelder (Flechsig 1990) individuell auszuwählen. Damit eine Individualisierung erfolgen kann, muss das Wissen für verschiedene didaktische Modelle aufbereitet werden.

Wollte man dies in Form von Lehrbüchern realisieren, wäre für jedes didaktische Modell ein dem Modell entsprechendes Lehrbuch zu verfassen. Das ist relativ aufwändig. Mit der Webdidaktik wird dieser Aufwand durch eine Analyse der in den didaktischen Modellen verwendeten Wissensarten reduziert. Viele Wissensarten (z.B. das Beispiel) werden in mehreren didaktischen Modellen verwendet. Es genügt daher, diese Wissensart nur einmal zu produzieren und sie dann im gewünschten didaktischen Modell einzusetzen.

Angesichts der Granularität der Wissenseinheiten im E-Learning wird dabei in jeder Wissenseinheit nur eine Wissensart repräsentiert. Zu einem Thema entstehen dabei mehrere Wissenseinheiten mit jeweils einer Wissensart. Die Wissenseinheiten zu einem Thema werden als *Lerneinheit* bezeichnet. Die Lerneinheiten stehen untereinander in einer sachlogischen Beziehung, die durch typisierte Relationen abgebildet wird.

Die manuelle Anordnung vorhandener Inhalte in ein didaktisches Modell ist nun immer noch zu aufwändig, um in der Praxis lernerspezifisch erfolgen zu können. Die Webdidaktik zielt daher darauf ab, diesen Prozess zu automatisieren. Dazu wird die didaktische Bedeutung der Wissenseinheiten von den Lehrenden zusätzlich zu den Wissenseinheiten eingegeben: Die Wissenseinheiten werden mit didaktischen Metadaten ausgezeichnet. Dafür wird das didaktische Metadatensystem (DML) verwendet.

Die DML stellt Metadaten für die Medientypen, die Wissensarten und die Relationstypen zwischen den Lerneinheiten bereit. Wissen wird auf dieser Grundlage in eine vernetzt-hierarchische Struktur gebracht. Die Hierarchie besteht aus zwei Ebenen:

- Wissenseinheiten, die durch ihren Medientyp und die Wissensart bestimmt sind und
- Lerneinheiten, die Wissenseinheiten zum gleichen Thema enthalten.

Die Vernetzung entsteht durch typisierte Relationen zwischen den Lerneinheiten.

Damit die Wissenseinheiten anhand der Metadaten automatisch anhand eines didaktischen Modells angeordnet werden können, müssen auf jeder Ebene didaktische Modelle spezifiziert werden. Auf der Ebene der Medientypen spezifiziert die Webdidaktik dazu Medienmodelle, auf der Ebene der Wissenseinheiten Mikromodelle und auf der Ebene der Lerneinheiten Makromodelle.

Die so entstehenden Lernpfade ermöglichen *fremdgesteuertes* Lernen, bei dem die didaktisch festgelegte Reihenfolge der Wissenseinheiten Schritt für Schritt abgegangen wird. Neben solchem fremdgesteuerten Lernen ist – schon um an nicht geplanten Orten der Lernsequenz entstehendem Wissensbedarf der Lernenden entsprechen zu können – ein *selbstgesteuertes* Lernen zu ermöglichen.

Kernproblem beim selbstgesteuerten Lernen ist die Frage, wie das gewünschte Wissen aufgefunden werden kann. Um benötigtes Wissen aufzufinden werden Retrievalverfahren verwendet (Buder 1991). Da die durch Relationen verbundenen thematisch bestimmten Lerneinheiten einen Thesaurus bilden, kann vorhandenes Wissen anhand des Thesaurus effektiv nachgewiesen werden (Wersig 1978). Aus dem Thesaurus können ein Index, eine Gliederung und topologische Wissenslandkarten erzeugt werden. Die Navigation innerhalb einer so gefundenen Lerneinheit erfolgt anhand der vorhandenen Wissensarten. Eine ergänzende Volltextsuche kann angeboten werden.

3 Die DML (Didactical Metadata Language)

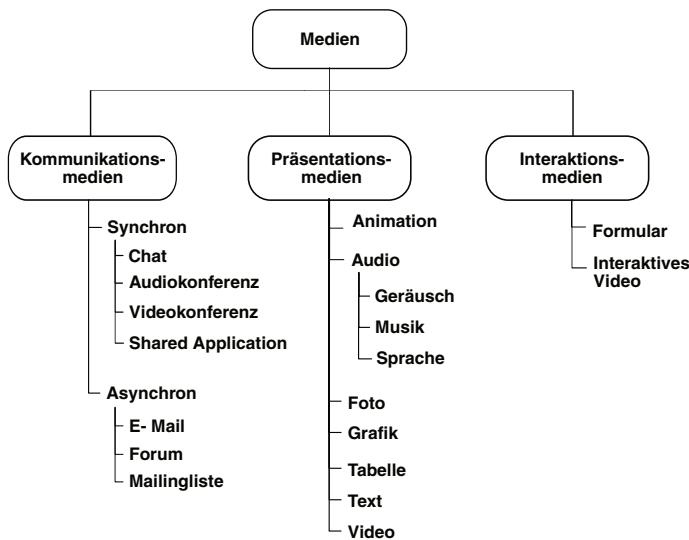
Mit der DML wird das didaktische Medium und die didaktische Wissensart einer Wissenseinheiten sowie die Relationstypen zwischen Lerneinheiten angegeben. Als grundlegende Arten werden dabei rezeptive, interaktive und kooperative Arten unterschieden.

- Rezeptive Arten bezeichnen passiv wahrgenommenes Wissen.
- Interaktive Arten bezeichnen Wissenseinheiten, bei denen Lernende in Interaktion mit dem Computer treten.
- Kooperative Arten bezeichnen Wissenseinheiten, in denen die Lernenden in Kommunikation mit Lehrenden oder Lernenden treten.

3.1 Medientypen

Eine Wissenseinheit kann in unterschiedlichen medialen Formaten präsentiert werden. Diese medialen Formate weisen unterschiedliche didaktische Qualitäten auf. Daher muss auch das mediale Format der Wissenseinheiten typisiert werden.

Die Medientypen werden unterschieden in Kommunikationsmedien, Präsentationsmedien und Interaktionsmedien. In den Unterkategorien geht es nicht darum, die technisch möglichen Variationen wiederzugeben, sondern darum, die Medien anhand der didaktischen Eigenschaften im Blick auf verschiedene didaktische Modelle zu typisieren. Die Kategorien der Präsentationsmedien erlauben auch Variationen des Abstraktionsgrads. Unterscheidet man zwischen bildlicher, ikonischer und symbolischer Darstellung ist klar, dass die bildliche Darstellung in einem Foto, die ikonische in einer Grafik und die symbolische in einem Text zu finden ist.



3.2 Mediale Verläufe

Die medial typisierten Wissenseinheiten können anhand des Medientyps in Lernpfade angeordnet werden. Als mediale Mikrostrategien schlägt Meder medial beschleunigte und medial verlangsamte Verläufe vor. Medial beschleunigt ist ein Verlauf, wenn die Rezeptionszeit pro Seite im Verlauf sinkt, medial verlangsamt ist ein Verlauf, wenn die Rezeptionszeit pro Seite im Verlauf steigt.

- Für einen medial beschleunigten Verlauf werden die Medientypen in folgender Reihenfolge angeordnet: Text, Klang, Grafik, Foto, Video.
- Bei medial verlangsamten Verlauf werden die Medientypen in folgender Reihenfolge angeordnet: Video, Grafik, Klang, Text.

Ergänzend zu diesen Verläufen können die Medien anhand des Abstraktionsgrades auch in abstrahierende und konkretisierende Verläufe angeordnet werden.

- Für einen abstrahierenden Verlauf werden die Medientypen in der Reihenfolge Video, Foto, Audio, Animation, Grafik, Tabelle, Text angeordnet.
- Für einen konkretisierenden Verlauf werden die Medientypen in der Reihenfolge Text, Tabelle, Grafik, Animation, Audio, Foto, Video angeordnet.

Die medialen Verläufe müssen dabei nicht notwendig vollständig sein. Die Reihenfolge ist so zu verstehen, dass die Medientypen in dieser Reihenfolge präsentiert werden, wenn verschiedene Medientypen vorhanden sind. Umgekehrt lässt sich aber fordern, dass verschiedene Medientypen produziert werden müssen, wenn die medialen Verläufe individuell variieren sollen.

3.3 Wissenseinheiten

Die Medientypen werden Wissenseinheiten zugeordnet. Eine Wissenseinheit kann verschiedene bildschirmgroße mediale Darstellungen einer Wissensart enthalten. Es wird nach den Grundkategorien zwischen rezeptivem, interaktivem und kooperativem Wissen unterschieden.

3.3.1 Rezeptive Wissenseinheiten

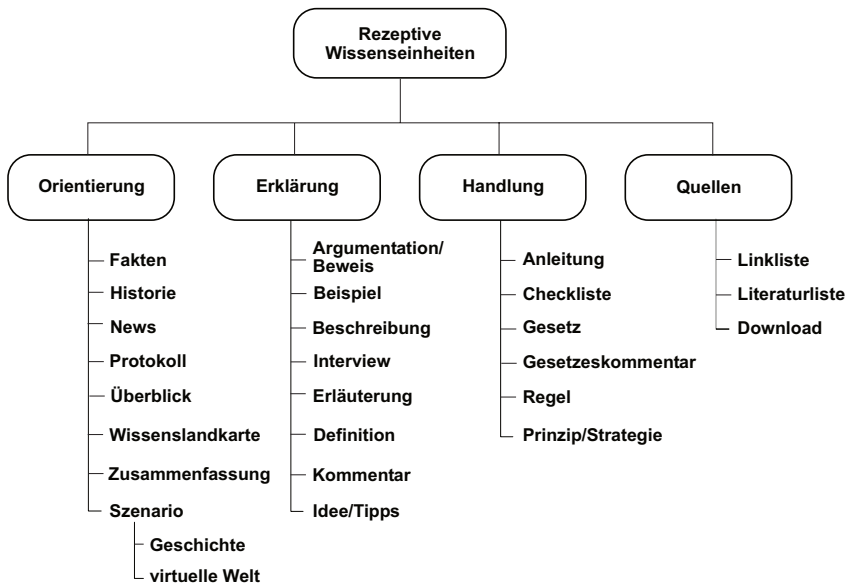
Die Wissensarten sind hierarchisch geordnet. Auf der ersten Ebene wird nach Flechsig (1990) Orientierungswissen, Erklärungswissen, Handlungswissen und Quellenwissen unterschieden:

- Orientierungswissen antwortet auf die Frage: „Was gibt es überhaupt?“
- Orientierungswissen wird benötigt, um sich in der Welt bzw. auf einem bestimmten Gebiet zurechtzufinden, ohne schon in spezifischer Weise erklären zu können und tätig zu werden. Es liefert eine erste Übersicht über das Thema und seinen weiteren Kontext. Es motiviert zur weiteren Entdeckung und weckt die Aufmerksamkeit. Dazu können Szenarien, Geschichten und Fakten etc. verwendet werden.
- Erklärungswissen antwortet auf die Frage: „Warum ist etwas so, wie es ist?“ Erklärungswissen liefert Argumente, mit denen Behauptungen und Empfehlungen belegt werden.
- Handlungswissen antwortet auf die Frage: „Wie ist es anwendbar?“ Handlungswissen bezieht sich auf reales Handeln von Menschen (Praktiken, Tech-

niken, Methoden und Strategien), auf ihr „Können“ und ihre Fertigkeiten („gewusst wie“).

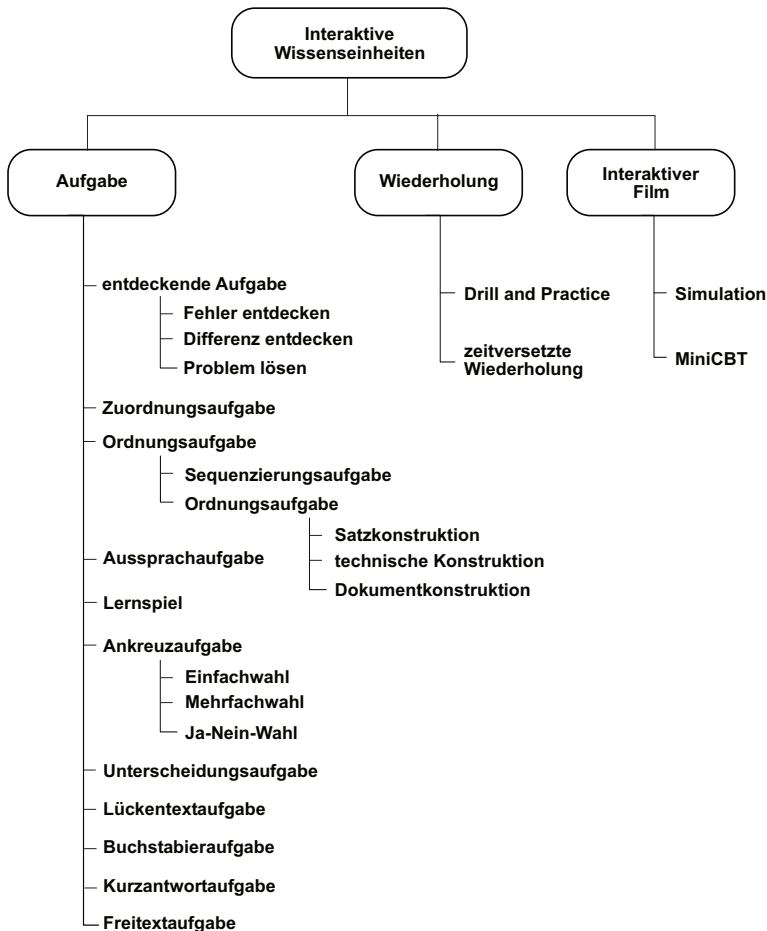
- Quellenwissen antwortet auf die Frage: „Wo gibt es weiteres Wissen?“. Es ist Wissen über Informationsquellen („gewusst wo“).

Die Unterteilung der rezeptiven Wissensarten erfolgt wie bei den anderen Wissensarten im Blick auf die didaktische Funktion des Wissens.



3.3.2 Interaktive Wissenseinheiten

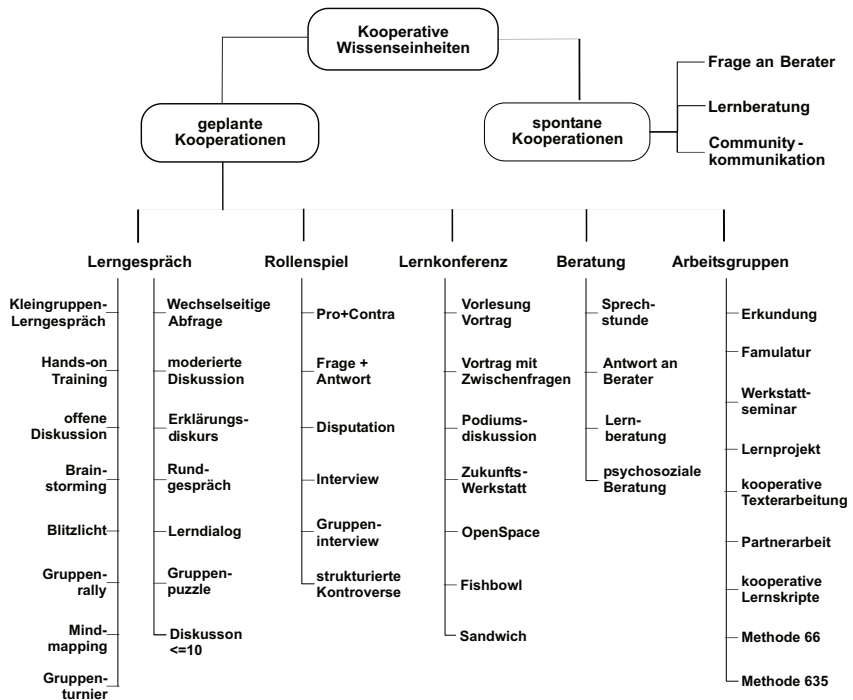
Die Aufgabentypen der interaktiven Wissensarten wurden im Rahmen des L-3-Projekts anhand einer Analyse der in computerbasierten Lernumgebungen verwendeten Aufgabentypen gewonnen. Sie beziehen sich nur auf Aufgaben, bei denen die Antworten durch Computertechnologie ausgewertet werden kann oder die Bearbeitung - wie z.B. oft bei entdeckenden Aufgaben - in Interaktion mit dem Computer erfolgt. Die Wiederholungsaufgaben sind vor allem für das Sprachlernen relevant.



3.3.3 Kooperative Wissenseinheiten

Ein Vorteil von Online-Lernumgebungen aus didaktischer Sicht ist die Möglichkeit, Kooperationen zu realisieren. Dazu genügt es nicht, lediglich die technischen Möglichkeiten anzubieten. Es ist erforderlich, Kooperationswerkzeuge anzubieten, die eine Strukturierung der Kooperation (Sozialform) ermöglichen und so zum einen in selbstgesteuerten Arrangements komplexe Abläufe unterstützen und zum anderen Lehrende in der effizienten Auswahl der gewünschten Kooperation unterstützen. Dabei können zugleich Basisfunktionalitäten verfügbar gehalten werden, um eine freiere Gestaltung von Kooperationen offen zu halten.

Zunächst ist aus didaktischer Sicht zwischen geplanten und spontanen Kooperationen zu unterscheiden. Spontane Kooperationen können jederzeit stattfinden. Geplante Kooperationen finden zu einem bestimmten Zeitpunkt oder an einer bestimmten Stelle im Lernverlauf statt. Die Systematik der kooperativen Wissenseinheiten wurde durch eine Analyse der didaktischen Literatur gewonnen. Diese erstellte Liste wurde systematisiert und im Blick auf das Online-Lernen auf Grundformen zurückgeführt. Die damit identifizierten Grundformen decken nicht alle beobachtbaren didaktischen Kooperationsformen ab, ermöglichen aber eine breite Variation.



Die Systematik der geplanten Kooperationen orientiert sich an der Gruppengröße. In Beratungen sind nur zwei Kooperationspartner anwesend, in Lerngesprächen, in Rollenspielen und Arbeitsgruppen sind kleine Gruppen (bis zu 30 TN) anwesend und Lernkonferenzen beziehen sich auf Großgruppen.

Die in der Abbildung genannten Kooperationsformen werden durch die Medienwahl spezifiziert. So kann der Lerndialog mit asynchronen Medien (z.B. E-Mail) und mit synchronen Medien (z.B. Chat oder Videokonferenz) durchgeführt werden. Die unterschiedlichen Rollen (Sprechen, Zuhören etc.), die Teil-

nehmerinnen- und Teilnehmerzahl und die Dauer der Kooperation müssen ebenfalls beim Anlegen der kooperativen Wissenseinheit spezifiziert werden. Dabei kann auch festgelegt werden, ob es sich um öffentliche oder private Kooperationsräume handelt. So können z.B. im Communitybereich private und öffentliche Kooperationsräume angelegt werden.

Geplante Kooperationen setzen in der Regel eine Aufgabe voraus, die in der Kooperation zu bearbeiten ist. Nach der Formulierung der Aufgabe ist eine Gruppenbildung erforderlich, die vor allem bei synchronen Kooperationen auch eine zeitliche Vorgabe einschließt. Anschließend kann die Kooperation durchgeführt und im letzten Schritt das Ergebnis gesichert werden.

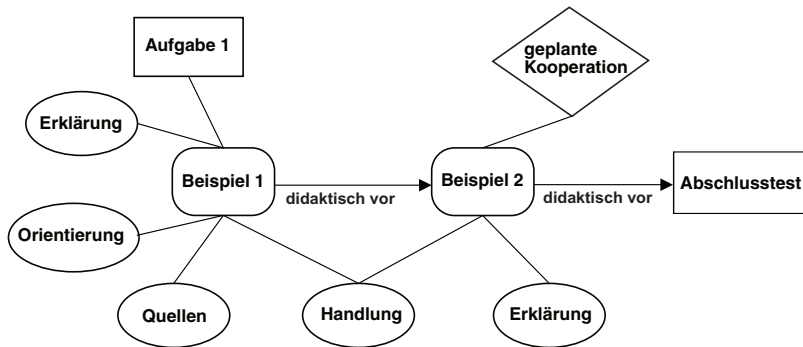
3.3.4 Mikrostrategien

Zwischen den Wissensarten einer Lerneinheit werden im DML-System didaktische Mikrostrategien bestimmt. Folgende Mikrostrategien sind spezifiziert: Klassisch theoriegeleitet, klassisch regelgeleitet, klassisch beispielgeleitet, klassisch nach Herbart, handlungsorientiert, aufgabenorientiert, problemorientiert, Überblickslernen, rollengeleitetes Lernen, testorientiertes Modell und das gruppenorientierte Modell.

Hier können aus Platzgründen nur zwei Mikrostrategien dargestellt werden. Der klassisch-theoriegeleitete Verlauf beginnt mit einer Orientierung, die Anschluss an Vorwissen enthält und angibt, worum es geht. Nach der Orientierung wird die Erklärung präsentiert, die sagt, warum die jeweilige Sache so ist, wie sie ist, und damit das theoretische Wissen enthält. Daran schließt sich eine Selbstkontrollaufgabe an. Nach der Selbstkontrollaufgabe wird Handlungswissen präsentiert, das sagt, was zu tun ist, wie also das begründete Wissen angewandt werden kann. Auch hieran schließt sich eine Selbstkontrollaufgabe an. Nach dem Abschlusstest wird mit Quellenwissen ein Verweis auf weitere Informationsquellen angeboten.



Die beispielorientierte Mikrostrategie gruppiert den Lernverlauf um eine Sequenz von Beispielen. Zu jedem Beispiel werden weitere Wissensarten angeboten, die die Lernenden je nach Bedarf abrufen können. Auch das beispielorientierte Modell wird mit einem Abschlusstest abgeschlossen.



Die Mikroverläufe bestimmen den didaktischen Verlauf innerhalb einer Lerneinheit. Zwischen den Lerneinheiten wird der Verlauf anhand der vorhandenen typisierten Relationen bestimmt.

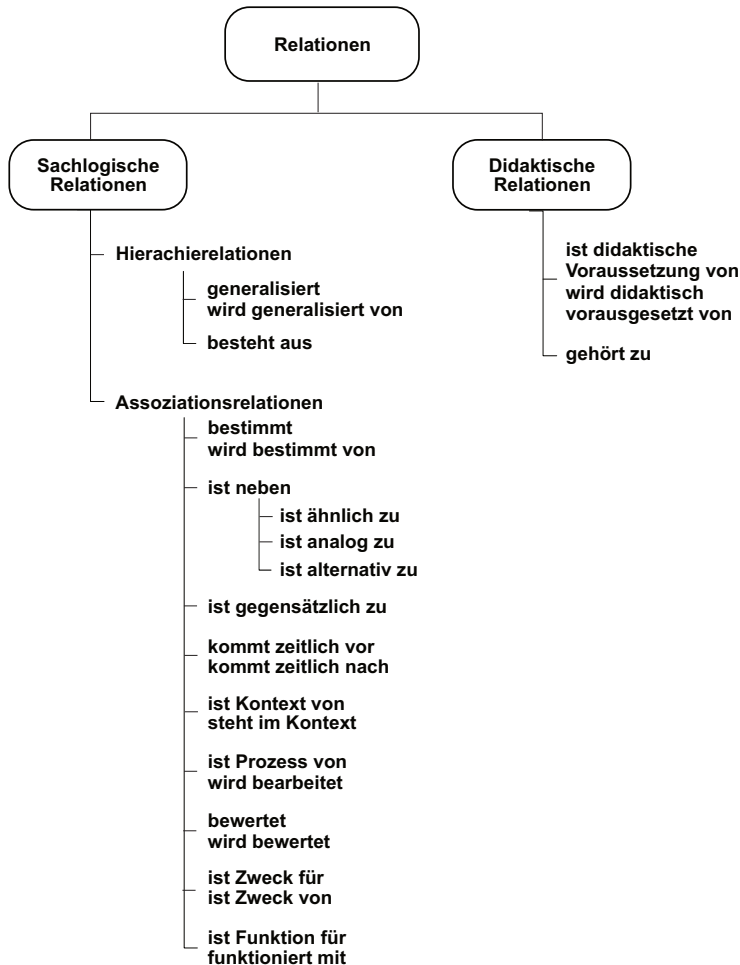
3.4 Relationales Wissen

Lerneinheiten umfassen Wissensseinheiten zum gleichen Thema. Die Lerneinheiten werden durch typisierte Relationen miteinander verbunden. Dadurch entsteht ein Thesaurus der in der Wissensbasis enthaltenen Begriffe. Auf die typisierten Relationen zwischen den Lerneinheiten werden didaktische Modelle abgebildet. Diese Modelle werden als Makrostrategien bezeichnet. Die typisierten Relationen dienen drei Zwecken:

1. Anhand der Relationen werden Lernpfade automatisch erzeugt.
2. Aus den Relationen werden Navigationshilfen, insbesondere Gliederungen, topologische Karten (Wissenslandkarten) und Verweise erzeugt.
3. Die didaktischen Relationen können verwendet werden, um bei der Zusammenstellung von Kursen durch Lernende Themen zu identifizieren, die während der Zusammenstellung des Kurses nicht berücksichtigt wurden, die aber notwendig enthalten sein müssen, damit die recherchierten Themen angeeignet werden können.

3.5 Relationstypen

In den Relationstypen wird zwischen sachlogischen und didaktischen Relationen unterschieden. Diese Unterscheidung ist notwendig, weil es didaktische Modelle gibt, die nicht ausschließlich auf die sachlogische Struktur abgebildet werden können. Anhand der typisierten Relationen können Lerneinheiten automatisch in didaktische Makrostrategien angeordnet werden.



3.6 Makrostrategien

Makrostrategien beziehen sich auf die Verläufe zwischen Lerneinheiten. In den Makrostrategien wird jeweils ein Relationstyp als dominant gesetzt. Dieser Relationstyp bestimmt im wesentlichen den Lernverlauf. Weitere Relationen werden herangezogen, um in Fällen, in denen der dominante Relationstyp keine Entscheidung ermöglicht, die Sequenz zu bestimmen. Die folgenden Strategien wurden von Meder entwickelt:

1. Goal-based bottom up: Für dieses zielorientiert-induktive Modell werden die Lerneinheiten anhand der Hierarchierelation (generalisiert/besteht aus)

angeordnet. Da dabei jede Lerneinheit mit mehreren hierarchisch höher geordneten Lerneinheiten verbunden sein kann, handelt es sich um einen Directed Acyclic Graph. Diese Struktur wird von unten nach oben zunächst so durchlaufen, daß jeweils nur das Orientierungswissen angegeben wird. Das Orientierungswissen der obersten Lerneinheit stellt das Ziel dar. Wenn dieses erreicht wird, werden die Lerneinheiten anhand der gewählten Mikrostrategie von unten durchlaufen. Dabei gibt es neben der Möglichkeit, während des ersten induktiven Durchlaufs alle Orientierungswissen zu präsentieren (next top), auch die Möglichkeit, zunächst nur das Orientierungswissen der obersten Lerneinheit anzuzeigen (top only) oder zunächst das Orientierungswissen der obersten Wissenseinheit anzuzeigen und anschließend die Hierarchie mit Anzeige der Orientierungswissen zu durchlaufen (top and next top).

In Kombination mit der problemorientierten Mikrostrategie realisiert diese Makrostrategie das problemorientierte Lernen. In Verbindung mit der aufgabenorientierten Microstrategie wird das aufgabenorientierte Lernen realisiert. Dabei kann zwischen wechselnden Aufgaben bei gleicher Rolle und durchgängig gleicher Aufgabe (Lernen in virtuellen Welten) unterschieden werden.

2. Bottom up: Diese Makrostrategie beschreibt ein rein induktives Lernen. Dabei werden wie bei der zielorientierten Strategie die Lerneinheiten anhand der hierarchischen Relationen geordnet. Die unterste Lerneinheit wird identifiziert. Wenn mehrere unterste Knoten vorhanden sind, wird die erste Lerneinheit anhand der „didaktisch vor“ - Relation ausgewählt. Anschließend werden die Lerneinheiten anhand der gewählten Microstrategie aufsteigen durchlaufen. Dabei können entweder zunächst die auf einer Ebene liegenden Lerneinheiten nach der „didaktisch-vor“ - Relation angezeigt werden, um dann in der Hierarchie aufzusteigen (Breite zuerst), oder es wird zunächst die Hierarchie bis zur obersten Lerneinheit durchlaufen, bevor die nächste Lerneinheit auf der untersten Ebene präsentiert wird.
3. Top down (deduktives Lernen): Bei dieser Makrostrategie werden die Lerneinheiten ebenfalls anhand der Hierarchierelationen angeordnet. Der oberste Knoten wird identifiziert und der Directed Acyclic Graph nach unten durchlaufen. Für die Entscheidungen bei Lerneinheiten, denen mehrere Lerneinheiten untergeordnet sind, gibt es zwei Varianten: Entweder wird der Graph erst bis zum untersten Punkt durchlaufen und dann der nächste Ast von oben begonnen, oder es werden erst die Lerneinheiten auf einer Ebene nacheinander durchlaufen und dann in die Tiefe verzweigt.
4. Das spiralmethodische Modell wird anhand der „ist Kontext von“ - Relation realisiert. Dabei kann entweder von einem Startpunkt aus immer weiter nach außen gegangen werden (dezentrativ), oder es wird ein Startpunkt außen ge-

- wählt, von dem aus dann immer weiter nach innen vorgegangen wird (konzentrativ).
5. Das aufbauende, konstruktive Modell (synthetisch) und das rekonstruktive Modell (analytisch) werden anhand der „Zweck-Mittel“- Relation und der „bestimmt“- Relation durchlaufen.
 6. Das dialektisch nach dem Urteilsquadrat verfahrenende Modell kombiniert Hierarchierelationen und Assoziationsrelationen so, dass die Lerneinheiten in der Anordnung des Urteilsquadrats der aristotelischen Logik durchlaufen werden.
 7. Die Netzwerkmethodete bietet die Lerneinheiten als Netzwerkstruktur an, so dass die Lernenden selbst entscheiden können, wie sie das Netzwerk durchlaufen. Dabei kann entweder eine Wissenslandkarte anhand der sachlogischen Struktur oder eine Wissenslandkarte anhand des Verlaufsnetzes eines didaktischen Modells angeboten werden.
 8. Guided Tour: Bei dieser Strategie werden die Lerneinheiten anhand der Relation „ist didaktische Voraussetzung von“ sequenziert. Die Lerneinheiten werden also unabhängig von der sachlogischen Struktur in einer festen Abfolge angeordnet.

Das hier vorgestellte Konzept zielt darauf ab, eine große Wissensbasis in kooperativer Autorenschaft zu erzeugen, in die individuelle Navigationen gelegt werden können. Dazu werden Inhalte aus der Wissensbasis zu Kursen zusammengestellt. Das kann entweder durch Lernende oder durch Lehrende geschehen. Anhand der Wissensseinheiten, Lerneinheiten und Relationen, die zwischen diesen ausgewählten Medientypen bestehen, kann dann ein Lernpfad automatisch generiert werden.

4 Praxisbeispiele

The screenshot shows a web browser window titled 'Lerndorf - Modalwert - Mozilla'. The address bar shows the URL: <http://www.lerndorf.uni-bielefeld.de/know/db/index.cgi?lexikon=index&navigation=>. The page has a green header with the 'lerndorf' logo. Below the header, there are navigation links: [Kurs beenden], [Modalwert], [Text], [Orientierung], [Handlungswissen], [Erklärung], [Beispiel], [Multiple Choice], [True / False], and [29/36]. A left sidebar contains a list of topics: Arithmetisches Mittel, Maße der zentralen Tendenz, Medianwert, Modalwert, and Summenzeichen. The main content area contains the following text:

In einer Untersuchung wurde die Frage gestellt, für welches Tier sich die Befragten entscheiden würden, wenn sie gezwungen wären, ein Tier eine Woche in ihrer Wohnung aufzunehmen. Die Codierung der Kategorien und die Häufigkeiten ihrer Nennung zeigt die folgende Tabelle:

	Hund	Katze	Hamster	Wellensittich	Goldfisch
Code	1	2	3	4	5
Häufigkeit	18	25	24	2	31

Die Mehrheit der Befragten würde einen Goldfisch aufnehmen. Der Modalwert ist also "Goldfisch".

In diesem Fall, wie bei allen nominalskalierten Daten, ist der Modus das einzig akzeptable Maß der zentralen Tendenz. Alle anderen Versuche einer Mittelwertbildung ergeben bei nominalskalierten Daten sinnlose Resultate.

© (Georg Hartl, 11.12.2003)

At the bottom of the page, there is a search bar with the text 'Such!' and a link '[Karte]'.

4.1 Forschungsprototyp

Ein Beispiel für die Umsetzung von Webdidaktik und DML zeigt der Bildschirmauszug aus dem Forschungsprototypen „Lerndorf“ in Abb. 8. Dargestellt wird eine Wissensseinheit mit der Wissensart „Beispiel“ aus der Lerneinheit „Modalwert“. Mit den Pfeilen oben rechts werden die Seiten des Kurses fremdgesteuert anhand des didaktisch bestimmten Lernpfads durchlaufen. Unten stehen mit einer Suche und mit topologischen Karten Retrievalwerkzeuge zur Verfügung. Navigationshilfen für das selbstgesteuerte Lernen werden mit den Links zu den Wissensarten des Themas oberhalb des Inhalts und der alphabetischen Themenliste an der linken Seite bereitgestellt. Die topologischen Karten, die alphabetische Themenliste, die Gliederung und die Links zu den Wissensarten werden aus den Metadaten generiert, die vom Autor bei der Erstellung der Wissensseinheit eingegeben worden sind.

4.2 Praxis der Produktion von Lernumgebungen

In der Praxis können nicht die benötigten Wissensseinheiten für alle hier vorgestellten didaktischen Modelle innerhalb einer Lernumgebung produziert werden. Es ist aber mit den vorgestellten Modellen möglich, zunächst die sachlogi-

sche Struktur zu entwerfen, um didaktische Relationen zu ergänzen und dann durch eine automatische Analyse festzustellen, welche der Modelle auch tatsächlich angeboten werden können.

Umgekehrt kann eine Lernumgebung gezielt für vorher ausgewählte didaktische Modelle produziert werden. Wenn z.B. aufgabenorientiert-zielorientiertes Lernen ermöglicht werden soll, ist klar, welche Wissenseinheiten mit welchen Wissensarten für jedes Thema benötigt werden und welche Relationstypen bei der Vernetzung zu verwenden sind. Dadurch kann die didaktische Komplexität des hier vorgestellten Konzepts im Einzelfall reduziert und für Autorinnen und Autoren handhabbarer gemacht sowie institutionellen Rahmenbedingungen (z.B. Finanzierung) angepasst werden. Das soll kurz am Beispiel der derzeit laufenden Entwicklung einer Online-Lernumgebung zur Einführung in die Statistik im Diplomstudiengang Pädagogik skizziert werden, aus der auch der Bildschirmauszug in Abb. 8 entnommen ist:

Eine Einführung in die Statistik im Diplomstudiengang Pädagogik wendet sich an eine Zielgruppe mit mathematischen Vorkenntnissen auf Abiturniveau. Ziel der Einführung ist es, zum einen ein Verständnis von statistischen Analysen zu erreichen, und zum anderen zur eigenen Anwendung von statistischen Verfahren, z.B. in Diplomarbeiten, zu befähigen. Ein Verständnis und die Fähigkeit zur eigenen Anwendung kann besonders gut mit beispielorientierten und handlungsorientierten Verläufen erreicht werden. Daher wurde die handlungsorientierte und die klassisch - beispielorientierte Mikrostrategie realisiert.

Für die Wahl der Makrostrategie ist ausschlaggebend, dass die Studierenden nur geringe Vorkenntnisse haben. Ohne Vorkenntnisse kann ein Gebiet jedoch kaum selbst strukturiert werden. Es gilt, auch eine Übersicht über das Gebiet der Statistik zu erreichen. Dafür sind verschiedene Makrostrategien geeignet. Besonders geeignet ist das zielorientiert-induktive Modell, mit dem zum einen das Ziel einen Verständnisrahmen anbietet, zum anderen die Struktur des Bereichs in einer einfachen hierarchischen Ordnung erschlossen werden kann. Für fortgeschrittene Studierende soll zusätzlich das Netzwerkmodell angeboten werden.

Das eng begrenzte Budget macht es unmöglich, unterschiedliche mediale Verläufe zu realisieren. Da für das Verständnis statistischer Analysen der Zusammenhang besonders relevant ist, wurde entschieden, das Orientierungswissen auch mit Animationen zu präsentieren. Daraus ergibt sich für jedes Thema folgende Wissensartenliste: Orientierungswissen (als Animation und als Text), Erklärung, Handlung, Beispiel, zwei Selbstkontrollaufgaben und Quellen (als Text). Die Themen müssen jeweils mindestens eine Ober- und eine Unterbegriffsrelation aufweisen und dort, wo das sachlich geboten ist, mit Assoziationsrelationen verbunden sein. Zu diesem Zweck wurde am Projektbeginn eine Gliederung der Themen in Form einer Wissenslandkarte entwickelt. Die hierarchischen Relationen ergeben sich unmittelbar aus der Wissenslandkarte. Zu-

sätzlich werden Relationen während der Inhaltsproduktion immer dann gesetzt, wenn in einer Wissenseinheit auf ein anderes Thema Bezug genommen wird. So setzt z.B. die Berechnung der Produkt-Moment-Korrelation die Kenntnis des arithmetischen Mittels voraus. Hierarchisch ist das arithmetische Mittel aber der univariaten Statistik zugeordnet und die Produkt-Moment-Korrelation der bivariaten Statistik. Hier ist daher eine zusätzliche Relation (ist Zweck für) zu setzen. Durch diese Relationen entsteht die geplante Netzstruktur.

4.3 Qualitätssicherung

Die Qualitätssicherung erfolgt zweistufig: Zunächst wird die didaktische Qualität geprüft. Dabei wird überprüft, ob der jeweilige Inhalt auch dem Thema und der Wissensart entspricht, ob also ein Beispiel für das arithmetische Mittel auch tatsächlich ein Beispiel für das arithmetische Mittel ist. Zugleich werden die gesetzten Relationen überprüft. Dazu wird kontrolliert, ob die erforderlichen Ober- und Unterbegriffsrelationen gesetzt sind und ob alle anderen Themen, die im Inhalt erwähnt werden, durch Relationen erschlossen sind. Dadurch kann zugleich die Kohärenz und Kohäsion der Wissenseinheiten sichergestellt werden. In einem zweiten Schritt werden die Wissenseinheiten inhaltlich geprüft. Dabei wird insbesondere die Verwendung der fachliche Terminologie und die Korrektheit der Formeln und Berechnungen überprüft.

Auf diese Weise kann mit der Webdidaktik auch mit geringen Mitteln eine didaktisch anspruchsvoll aufbereitete Online-Lernumgebung erzeugt werden.

6 Literatur

- Buder, Marianne (Hrsg.): Grundlagen der praktischen Information und Dokumentation. 3. neu gefasste Ausg., München.
- Flechsig, Karl-Heinz (1983): Der Göttinger Katalog didaktischer Modelle. Nörten-Hardenberg.
- Flechsig, Karl-Heinz (1990): Einführung in CEDID. Göttingen.
- Flechsig, Karl-Heinz (1990): CEDID. Ein wissensbasiertes System für computerergänztes didaktisches Arbeiten. Göttingen.
- Heimann, Paul (1976): Didaktik als Unterrichtswissenschaft. Klett.
- Höningwald, Richard (1927): Über die Grundlagen der Pädagogik. München.
- Meder, Norbert (1996): Der Sprachspieler. Ein Bildungskonzept für die Informationsgesellschaft. In: Vjschr. f. wiss. Päd. (2) 1996, S. 145-162
- Meder, Norbert (1998): Neue Technologien und Erziehung/Bildung. In: Borrelli, M.; Ruhloff, J.: Deutsche Gegenwartspädagogik Bd.III, Hohengehren, S. 26-40

- Meder, Norbert (2000a): Evaluation von Lern- und Spiele-Software. In: Fromme, J.; Meder, N.; Vollmer, N.: Computerspiele in der Kinderkultur. Opladen, S. 176-227.
- Meder, Norbert (2000b): Wissen und Bildung im Internet - in der Tiefe des semantischen Raumes. In: Marotzki, W.; Meister, D.; Sander, U.: Zum Bildungswert des Internet. Opladen, S. 33-56
- Meder, Norbert (2003): Didaktische Anforderungen an Lernumgebungen: Die Web-Didaktik von L3. In: Ehler, U. D. (u.a.): E-Learning Services im Spannungsfeld von Pädagogik, Ökonomie und Technologie. Bielefeld, S. 50-69.
- Porac, J.F.; Howard, T. (2002): Managing Cognition and Strategy. Issues, Trends and Future Directions. In: Pettigrew; Howard; Whittington (Eds.): Handbook of Strategy and Management. London.
- Swertz, Christian (2000): Ausbildung zum Gebrauch didaktischer Ontologien. In: Ohly, H.P.; Rahmstorf, G.; Sigel, A.: Globalisierung und Wissensorganisation. Würzburg, S. 431-442
- Swertz, Christian (2001): Computer und Bildung
(<http://archiv.ub.uni-bielefeld.de/diss/2001/0033/diss.ps> [16.7.2002]).
- Swertz, Christian (2002): Konzepte und Methoden zur Qualitätssicherung bei der Produktion von hypertextuellen Online-Lernumgebungen. In: Medien-Pädagogik (1) (<http://www.medienpaed.com/02-1/swertz1.pdf> [15.4.2002])
- Walsh, J.P. (1995): Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane. In: Organizational Science 6 (3), 280-321.
- Wersig, Gernot (1978): Thesaurus-Leitfaden. München.

The C₂M project: a wrapper generator for chemistry and biology

Paul van der Vet, Eelco Mossel

Department of Computer Science, University of Twente
P.O. Box 217, 7500 AE Enschede, the Netherlands
Phone +31 53 489 3694, fax +31 53 489 3503
e-mail {vet,mossel}@cs.utwente.nl

Modern science relies on the availability of resources accessible over the web. Each resource uses its own format, among other things because science is highly dynamic and tasks change frequently. In other words, format multiplicity is a fact of life. Data interoperability relies on the presence of wrappers. The C₂M project aims to build a system that supports quick and easy generation of lightweight wrappers by providing a language in which formats can be specified. Because the project was originally aimed at chemical applications, the name “C₂M” is a chemical formula-like abbreviation of “chemical configurable middleware”. For reasons of exposition, we will be looking at a simple chemical format in this paper. C₂M can be successfully applied in other domains as well. The C₂M language has been designed to be easy to learn and use, yet it is sufficiently formal to allow unambiguous description of formats. There is a provision for including documentation, and in fact specification writers are encouraged to do so copiously. Underlying the design of the C₂M language is the intuition that there is such a thing as language ergonomics.

1 Introduction

The C₂M project¹ is concerned with the development of a user-friendly programming language dedicated to producing wrappers. A wrapper is a piece of software that interconverts between different data formats. The project has started by concentrating on (bio)chemical formats as examples, but the language is suited for other disciplines as well. The problem of format interconversion is encountered in every project in which co-operation between multiple, heterogeneous resources is required. Data tend to come in formats that do not meet the requirements of the project at hand. Data coming from different sources almost al-

1 P.E. van der Vet, H.E. Roosendaal, and P.A.T.M. Geurts, “C₂M: configurable chemical middleware”, *Comparative and Functional Genomics* 2 (2001), 371—375.

ways come in different formats. Co-operation between resources can only be achieved by interconversion. If there is a need to dynamically add resources to those already used, some internal uniform data format has to be chosen to ensure smooth interoperability. A pair of wrappers connects every resource to the user desktop (figure 1).

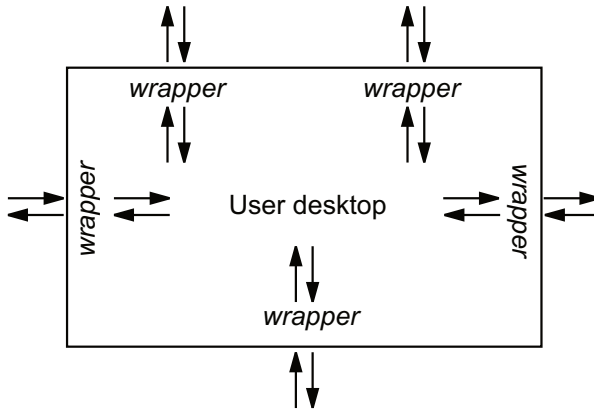


Figure 1 Providing the user desktop with access to multiple, heterogeneous resources

The construction of wrappers has received enormous attention in the past decades. CORBA² is designed to be the ultimate solution to the problem of resource interoperability. A CORBA architecture is robust, portable across many platforms, and scalable. In practice, however, CORBA tends to be heavy-weight and static. Developing an interface in CORBA's IDL language is laborious. IDL code has procedural flavour and semantics of data are implicit. Perhaps for these reasons, the advent of CORBA has not stopped researchers from developing other approaches to resource interoperability.³

Turning to the more specific issue of wrappers, many software packages come with built-in wrappers, but their range is restricted and including a new format is either impossible or very laborious. The other, very common alternative is to have project members write wrappers themselves. We will call these "Roll your own" (RYO) wrappers. RYO wrappers are typically written in procedural languages such as awk, Java, Perl, Python, or just straight C. Literature about the influence of notations and formalisms on human performance in a spe-

² <http://www.corba.org>

³ See, for example, S. Bergamaschi, S. Castano, M. Vincini, and D. Beneventano, "Semantic integration of heterogeneous information sources", *Data and Knowledge Engineering* 36 (2001), 215—249, and C.A. Goble et al., "Transparent access to multiple bioinformatics information sources", *IBM Systems Journal* 40 (2001), 532—551.

cification task, however, suggests a preference for languages that describe constraints on solutions rather than procedures for finding solutions.⁴ In the wrapper domain, this preference translates into one for wrappers built from high-level, declarative descriptions of the formats involved. Indeed, RYO wrappers in a procedural language are generally hard to read and maintain, particularly so over intervals of a year or more. CORBA's IDL objects suffer from the same defect. On the positive side, RYO wrappers present a flexible, lightweight approach to resource interoperability, wholly in line with the dynamic nature of science on the Web. The way out is to use wrapper generators. A wrapper generator generates wrappers from high-level descriptions. Because implementation details are hidden from view, wrapper generators facilitate the task of building and maintaining RYO wrappers. The main fundamentals of wrapper generators are known from compiler building.⁵ An approach to wrapper generation has been published fourteen years ago by Mamrak and co-workers.⁶ Since then, many approaches to wrapper generation have been published. The WWW Wrapper Factory (W4F) approach⁷ resembles C₂M but is largely confined to XML formats. The C₂M language bears resemblance to the idea of token-templates.⁸ Automated wrapper generation⁹ and wrapper verification¹⁰ have also received attention.

C₂M builds on these experiences to present a wrapper generator that facilitates the design and maintenance of RYO wrappers. It is not intended to supplant existing software for regular conversion jobs. Rather, C₂M supports irregular jobs such as extracting a table with protein-protein interactions from the

4 For example, B.W. van Schooten, Development and specification of virtual environments, Ph.D. thesis, Parlevink group, University of Twente, Enschede, the Netherlands, 2003, chapter 5; also available as <http://wwwhome.cs.utwente.nl/~schooten/proefschrift.pdf>; B. Khazaei and C. Roast, "The usability of formal specification representations", in: G. Kadoda (ed.), Proceedings of the 13th Workshop of the Psychology of Programming Interest Group, Bournemouth UK, April 2001, pp. 305—310.

5 The classic text on compiler building is A.V. Aho and J.D. Ullman, Principles of compiler design, Reading MA: Addison-Wesley, 1979.

6 S.A. Mamrak, M.J. Kaelbling, C.K. Nicholas, and M. Share, "Chameleon: a system for solving the data-translation problem", IEEE Transactions on Software Engineering, 15 (1989), 1090—1108; S.A. Mamrak, C.S. O'Connell, and J. Barnes, The integrated Chameleon architectures, Englewood Cliffs NJ: Prentice Hall, 1994.

7 A. Sahuguet and F. Azevant, "Building intelligent Web applications using lightweight wrappers", Data and Knowledge Engineering 36 (2001), 283—316.

8 B. Thomas, "Token-templates and logic programs for intelligent web search", Journal of Intelligent Information Systems 14 (2000), 241—261.

9 N. Kushmerick, "Wrapper induction: efficiency and expressiveness", Artificial Intelligence 118 (2000), 15—68; P.B. Golgher, A.H.F. Laender, A.S. da Silva, and B. Ribeiro-Neto, "An example-based environment for wrapper generation", in: S.W. Liddle, H.C. Mayr, and B. Thalheim (eds.), Conceptual Modeling for E-Business and the Web, Berlin: Springer, 2000, pp. 152—164.

10 N. Kushmerick, "Wrapper verification", World Wide Web 3 (2000), 79—94.

XHTML version of a journal article, extracting certain numbers from a file and output them for SPSS processing, turn metadata into a uniform format for the MPRESS project,¹¹ or, indeed, turn a high-level description of a resource into a piece of CORBA IDL code. Although originally intended for conversion of plaintext file formats, the design supports the conversion of any format because it can handle Unicode character sets and control codes; see also section .

2 An example: CT files

The C₂M language will be discussed by means of an example. A full description of the language will be given in the manual, which is under construction. The running example is the CT (connection table) format as exported by the ChemDraw package for drawing molecular structures. The CT format is one of the many available formats for representing chemical structure.¹² It is a very simple format; space restrictions forbid a more realistic example. A CT file just lists the graph: the atoms and the bonds between them. As usual in such files, all hydrogen atoms and the bonds connecting hydrogen atoms to other atoms are left out. The CT file for a popular molecule, CH₃CH₂OH or ethanol, is given in figure 2. The first line lists the filename but may contain any string. The second line lists the number of atoms and the number of bonds, respectively. Their presence dates from the Fortran era, when the reading program had to know in advance how many lines to read.

```
ethanol.ct
  3    2
    -0.8667  -0.2500  0.0000  C
      0.0000   0.2500  0.0000  C
      0.8667  -0.2500  0.0000  O
  1  2  1  1
  2  3  1  1
```

Figure 2 CT file for the ethanol molecule

The next three lines provide atom information, one atom per line. The three numbers in front are co-ordinates used by ChemDraw to reconstruct the drawing from which the file has been generated. The presence of such co-ordinates in chemical structure files is quite common. At the end of each atom line we find

¹¹ J. Plümer, “MPRESS – Transformation von Metadaten Formaten”, This volume.

¹² T. Engel and J. Gasteiger, “Chemical structure representation for information exchange”, Online Information Review 26 (2002), 139—145.

the chemical symbol. Implicitly each of the atoms is numbered, starting with 1. This becomes evident in the last two lines of the file, which list bonds. Each has four numbers: the first two are the numbers of the atoms the bond connects. Then follows the nature of the bond (single, double, ...). The last number, called a bond order by ChemDraw, again is a drawing aid for ChemDraw.

3 The C₂M language and wrapper generator

3.1 Introduction

The C₂M project has concentrated from the start on the language in which the wrapper task is specified. The present section attempts to provide an impression of the C₂M language. It has to fulfill contradictory requirements: ease of use, sufficient freedom to express many formats, and sufficient formality to express unambiguous format descriptions. Because existing format specifications are often sloppy and imprecise, we consider it mandatory that the language can be used to publish and share format specifications even if no wrapper is intended. This gives the language a distinct advantage as communication medium over procedural languages, because in procedural languages format specifications are by necessity implicit in the procedures. In addition, there has to be a compiler that can produce an executable wrapper from format specifications. Finally, it would be nice if the wrapper runs in acceptable time and exhibits agreeable scaling behaviour. The result, no matter how we proceed, is inevitably a compromise.

As we have explained elsewhere (ref.), C₂M conversion proceeds over an intermediate format. Having an intermediate format reduces the number of converters required from $n^2 - n$ to $2n$, which already pays off for $n > 3$. Also, it prepares for middleware tasks such as comparison of two or more sources and merging of sources. When C₂M has read data from a file, it stores these data in a format called C₂M's *native representation*. An ontology serves as a template for the native representation. Ontologies are specified by the user and can be tailored to the task at hand. Conversion of data from a source format *A* into a target format *B* thus proceeds in two steps. The first step takes its information from the specification of format *A* and an ontology as input. The specification of format *A* includes the specification of the accompanying ontology. The second step takes its information from the specification of format *B*. See figure 3.

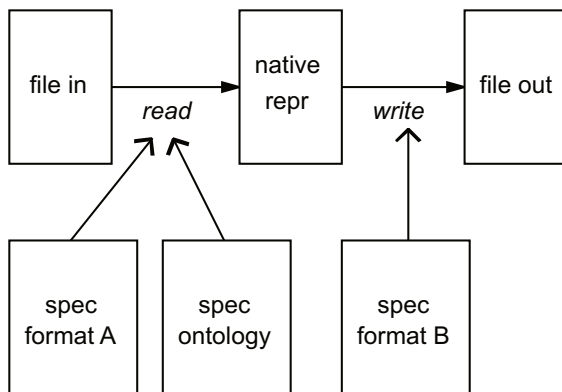


Figure 3 File conversion by the C₂M system

C₂M views any file as a string of characters. The string can be analysed as consisting of substrings. These substrings fall into one of three categories: meaningful strings, landmarks, and redundant strings. The decision to view a particular string as meaningful or redundant depends somewhat on the application. *Meaningful strings* represent the information we want to read from the source file or write to the target file. In the CT example file of figure 2, for example, the meaningful strings are, on the atom lines, the co-ordinates, element symbols; and on the bond lines, the atom numbers, and the bond types. This means that we view the entire first and second lines and the ChemDraw bond order as *redundant strings*. The space characters in front of every line except the first are likewise redundant. *Landmarks* serve to delimit and identify records and fields. In the CT file, the space characters between data on the same line and the end-of-line characters serve as landmarks. In other formats, fields are identified by special strings. In the Unix refer bibliography format, for example, a two-letter string at the start of each field serves as field identifier.

The C₂M task is to extract the meaningful strings from the source file while preserving their mutual relations like order, and to store those strings, possibly after some modification, as the native representation. Modifications include simple calculations and character transformations; also, a table can be specified that for each input gives the corresponding native representation. To do this, we have to name those strings and the larger structures of which they form part. This can be done by parsing the source file. Next, the meaningful strings are extracted and possibly converted into the format required by the native representation according to conversion rules specified by the user. The conversion rules are known as semantic bindings. The parsers are generated from user-supplied grammars (compare, for example, ref.). This enforces a natural separation bet-

ween form and content: the syntax of the source file is covered by the grammar while the semantics of meaningful strings are fixed by the conversion rules.

The overall view of the system is shown in figure 4, see ref. 1 for a more extensive discussion. The process starts with ontology and file format specifications supplied by end-users and/or content providers. The C₂M system consists of the C₂M compiler and a converter core. The C₂M compiler turns each specification into source code in some programming language, or into object code. The appropriate compiler and/or linker is then used to produce an executable that is able to convert from and into each format for which a specification is provided. Another program called the *documenter* turns the same specifications into XML or LaTeX sources, to produce documents for human consumption by the appropriate renderers. The documentation consists of the comments included in the specification and the specification itself. The principle followed here is that of literate programming:¹³ always derive code and documentation from the same source.

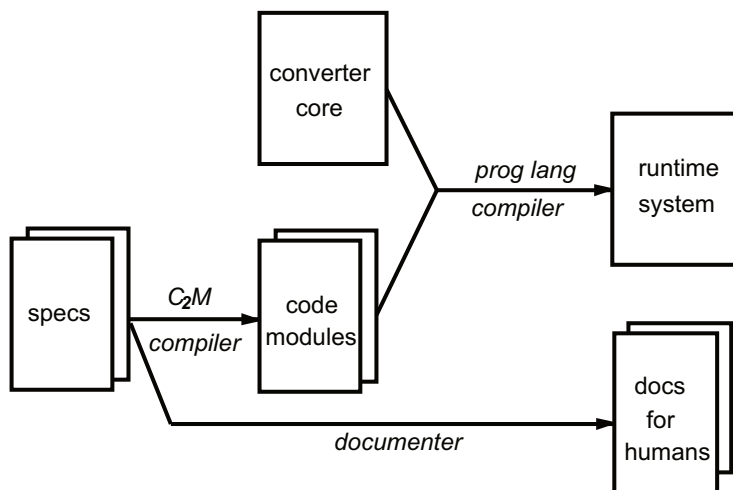


Figure 4 Operating the C₂M system

In the following subsection, we will discuss a number of features of the C₂M language. A simple ontology of molecules and the format specification of CT files are given as appendices.

13 D.E. Knuth, *Literate programming*, Palo Alto CA: Center for the Study of Language and Information of Stanford University, 1992.

3.2 Basic structure of a C₂M specification

Ontology and file format specifications are written by the user to determine the way C₂M converts one file into another. Each specification comes in a file of its own. It includes both the specification itself and documentation to explain the choices made. A specification file is a plaintext file except that literal strings may consist of any character provided it is taken from some pre-defined character set such as Unicode.¹⁴ The precise definition of a plaintext file is beyond the scope of the present paper; it roughly corresponds to the Unicode C0 character set¹⁵ from which the control codes have been removed except for end-of-line. The plaintext requirement currently also applies to the documentation included in the file; this constraint will be relaxed in future versions.

The leading idea in designing the C₂M language has been to comply as much as sensible with well-known syntax practices without compromising the efficiency of the code generator. This has resulted in modest use of tags with or without attributes, as known from HTML syntax, and the use of well-known operators or their plaintext imitations, such as “::=” or “->” (imitation of “→”) for grammar rules, “<-” (imitation of “←”) for instantiation rules, and “=^” (imitation of “≐”) for correspondence rules (the nature of these rules will be explained below).

Below, we will use the Arial font to refer to characters and strings that have to be present literally in the specification, and the italic Arial font is used to refer to variables that have to be replaced by literal strings in the final specification.

Every C₂M specification has the general structure:

```
<C2M-SPECIFICATION name=name
                        type=type>
```

```
blocks
```

```
</C2M-SPECIFICATION>
```

where *type* currently can be one of the two ontology or file-format; *name* is chosen by the user and is required to be unique within each C₂M implementation. The *blocks* are also enclosed in tags. What the blocks are depends on the nature of the specification: *ontology* or *file-format*.

Documentation can be included wherever a tag or rule is expected, but never within a tag or rule. Documentation is enclosed in tags <TEXT> and </TEXT> (case-insensitive). Specification writers are encouraged to insert copious documentation in their files.

¹⁴ The Unicode Consortium, The Unicode standard, version 3.0, Reading MA: Addison-Wesley, 2000.

¹⁵ Ref. , pp. 336—340.

3.3 Ontology specifications

Ontology specifications consist of a single block. The ontology formalism has been kept to an absolute minimum: it is expressed in a frame language in which concepts can have attributes. The relations between concepts and attributes are left unspecified. A simple ontology of molecules is used as illustration.

Concepts come in three sorts. *Complex concepts* have an arbitrary number of unique attributes. Attributes are concepts, and thus may have attributes themselves. An example is the concept of atom:

```
atom = chemical-symbol, id, dr-coord-x, dr-coord-y, dr-coord-z
```

where id is the number that identifies the atom in the molecule. The attributes are delimited by a comma followed by a space. In fact, any combination of comma's, space characters, and end-of-line characters, provided there is at least one of them, can be used to delimit tokens in a specification file.

Concepts with a repeated attribute have a single attribute that can, however, occur an arbitrary number of times. For example:

```
atom-list = repeated( atom )
```

In the native representation, the atom attribute will occur as many times as dictated by the source file: three times in the CT example of figure 2.

Primitive concepts have no attributes. They can hold an instantiation, normally a string read from a source file. In the rule above for atom, the concept chemical-symbol is primitive. If the CT file of figure 2 is read, it will hold the string "C" or "O", depending on from which line the instantiation is derived.

3.4 File format specifications

File format specifications consist of two or four blocks, depending on whether the specification is intended for reading or writing only (two blocks) or for both reading and writing (four blocks). The read and write task each are specified in a syntax block that contains a grammar and a semantics block that contains so-called semantic bindings. Each block is enclosed in tags.

In specifications for both reading and writing, it may seem wasteful to specify the grammar and the semantic bindings for reading and for writing separately. We can only use the same grammar for reading and writing when it is reversible. A grammar is reversible when there is no syntactic indeterminacy. Unfortunately, syntactic indeterminacy in the kind of files addressed by C₂M is common. For example, in the CT file of figure 2, every line save the first starts with space characters, but whether they are there and, if so, how many are there does not matter. In file types with explicit field identifiers, field order within a record can be free. The grammar for reading allows for indeterminacy in these cases. The

grammar for writing, however, needs explicit instructions that rule out syntactic indeterminacy. Semantic bindings, by contrast, can be reused to some extent. To be fit for reading and writing, the particular rule in the semantic bindings has to be bijective. In that case, the rule has to be specified in the semantic bindings for reading only.

3.5 File format specification files: reading

The grammar for reading follows the rules for BNF (Backus-Naur Form) grammars. A rule that describes any CT file in its entirety is:

```
ct-file -> ct-first-line ct-numbers-line ct-atom-lines ct-bond-lines
```

where the prefix `ct` is inserted to remind the human reader of the particular format we are describing. The four symbols at the right-hand side have to be defined elsewhere in the grammar such that ultimately we arrive at the level of strings. `C2M` has a number of built-in character and string classes, for example the character classes `letter`, `digit`, and `end-of-line`; and the string classes `word` (only letters), `spaces` (any sequence of space characters), `string` (any sequence of non-space characters), `line` (any sequence of characters except `end-of-line`), `integer`, and `float`. Using these, we may write for example:

```
ct-first-line -> line end-of-line
```

```
ct-atom-line -> spaces ct-coord-x spaces ct-coord-y spaces
               ct-coord-z spaces ct-chem-symbol end-of-line
```

```
ct-coord-x -> float
[etc.]
```

The symbol `ct-atom-lines` stands for the series of consecutive lines with information on the atoms that constitute the molecule. Because the number of those lines is mentioned in the second line of the CT file, we could write a grammar rule that makes use of this information. This would be laborious, however; it is easier to write:

```
ct-atom-lines -> ct-atom-line+
```

where the BNF-operator `+` signals one or more occurrences. A rule of this kind always explicitly records the position information so that, in this case, we can identify the *n*th occurrence of `ct-atom-line` in the list `ct-atom-lines`.

At this point, the substrings of a CT file and groupings of those substrings have been named. We can rely on the parser to build a parse tree as it analyses the source file from which we can extract groupings and substrings by name. We now have to tell `C2M` which meaningful strings are used to instantiate primitive concepts of the ontology and how to translate them in order to comply with the

requirements of the native representation. This has to be done at two levels. First, we have to specify correspondences between groupings, for example:

```
atom-list =^ ct-atom-lines
```

Next, we have to specify conversions, if any. This is done by means of instantiation rules. We decide we want to store the strings as they are, for example:

```
chemical-symbol <- ct-chem-symb
```

This does not give us a way to instantiate the id concept, because there is no meaningful string in the source file that provides the required information. Instead, we make use of the fact that grammar rules for repeated occurrences record position information:

```
id <- position(ct-atom-line, ct-atom-lines)
```

which translates roughly as “for each atom, instantiate the id concept with the number that records the position of ct-atom-line in the list ct-atom-lines”.

Instantiation rules can be left out of the format specification. The effect is that the corresponding primitive concept in the ontology will not be instantiated, which in turn means that the entire concept will be absent from the native representation. This feature can be used to incorporate a weak form of disjunction in ontologies. To promote reuse of ontologies, one can add all primitive concepts that might be needed. The decision which concepts to instantiate is taken in the file format specification.

3.6 File format specification files: writing

The grammar for writing is much like the grammar for reading, except that indeterminacy is not allowed. This means, among other things, that pre-defined string classes such as spaces or line cannot be used. In writing, every grammar symbol will have to produce output. In addition to a grammar, a format specification for writing has semantic bindings (correspondences and instantiations) to translate the native representation into the target format. If a format specification is used for both reading and writing, bijective semantic bindings already specified in the read part do not have to be repeated in the write part. Correspondences are always bijective.

We could in principle reuse the first rule of the grammar of a CT file for reading:

```
ct-file -> ct-first-line ct-numbers-line ct-atom-lines ct-bond-lines
```

but we cannot use other read rules for writing. In the first place, we have to say what should come at the first line, for example:

ct-first-line -> "File generated by C₂M" end-of-line

which puts the string at the right-hand side on the first line of the target file and ends it with an end-of-line character. Next, at the second line called ct-numbers-line, the numbers of atom lines and bond lines have to be filled in. In the native representation, the numbers of atoms and bonds are not present explicitly. Therefore we have to establish them in another way. The C₂M language has the construct:

fl-text(Supersymbol, Symbol, NrOcs)

that is a canned version of the grammar rule

Supersymbol -> Symbol+

except that the number of occurrences of Symbol is fixed to be NrOcs. This can be either an explicit number or a grammar symbol that writes that number at the appropriate place. In the latter case, there are two occurrences of the NrOcs symbol that should be at the right-hand side of the same grammar rule. Using this construct, we may choose to write the entire CT file using the rule:

ct-file -> "File generated by C₂M" end-of-line

space space nr-atom-lines space space nr-bond-lines end-of-line

fl-text(ct-atom-lines, ct-atom-line, nr-atom-lines)

fl-text(ct-bond-lines, ct-bond-line, nr-bond-lines)

which will write a correct CT file provided there are correct rules for ct-atom-line and ct-bond-line elsewhere in the grammar for reading.

4 Implementation

The current version of C₂M is implemented in Prolog.¹⁶ We believe this has considerably eased the development of the system, but we also want to stress that the same functionality can be realised in any other programming language. Indeed, sharing and reuse of C₂M specifications should be independent of the programming language or languages used.

The implementation has been inspired by work in natural-language processing. For example, the two read steps are syntactic and semantic analysis, precisely as in many natural-language understanding systems. For syntactic analysis, we use the parser generator built into Prolog. Semantic analysis is based on knowledge-based system technology. The converter core of C₂M incorporates a

¹⁶ L. Sterling and E. Shapiro, *The art of Prolog*, Cambridge MA: MIT Press, 1994 (2nd edition); I. Bratko, *Prolog programming for artificial intelligence*, Harlow: Addison-Wesley, 2001 (3rd edition).

special-purpose inference engine that uses the semantic parts of a file format specification as knowledge base.

A few points in the implementation merit attention. Using grammars in specification files to generate parsers is a very general way to handle the syntax of formats. Because tokens can be delimited by any character, a separate tokenisation step is not foreseen. Because the parser analyses the source file on a character-by-character basis, C₂M can handle any type of file but conversion will proceed slower than in the presence of a tokenisation step. Even control codes can be defined as characters and handled appropriately. Molecular structure files for not too large molecules (say, with 1000 atoms or less) are processed within a second. Preliminary experiments have shown that the scaling behaviour of the system approaches linear behaviour.

The C₂M compiler in fact converts a specification file into a Prolog file. In other words, the compile task can in principle also be done by the C₂M converter, given the appropriate specification files. One reason not to do so is efficiency. In specification files we do know what the token delimiters are, and therefore we can insert a tokenisation step before the parsing step. Also, the C₂M compiler optimises to some extent. Code optimisation is beyond the capabilities of a format converter such as C₂M.

Finally, the Prolog implementation we used is Quintus Prolog¹⁷ because it is robust and very fast. Although most constructs are standard Prolog,¹⁸ we sometimes made use of Quintus built-ins because they are more efficient. Quintus Prolog is proprietary software. We are currently contemplating to implement future versions in Java.

5 Conclusion

We have given a very cursory overview of the C₂M language. The language is designed with ease of use as primary concern, where “ease of use” has been operationalised mainly as “familiar” in a number of ways. User surveys will have to bear out whether we succeeded in this respect. A wide scope has been another design criterion. It has been operationalised by analysing foreign files at the level of individual characters. C₂M, however, does not aim to make existing software superfluous. For example, the provision of extensive options for calculations has not been considered because there is well-known and widely available software that can do this. A desktop system built around C₂M will perform calculations by having C₂M convert some intermediate result into a query to a calculation package, and having C₂M convert the package’s output into a suitable form for further processing.

17 <http://www.sics.se/isl/quintus/>

18 P. Deransart, A.A. Ed-Dbali, and L. Cervoni, *Prolog: the standard*, Berlin: Springer, 1996.

C₂M is now tested on a large number of formats to detect weaknesses and omissions. Future work will be directed at an extension of the possibilities of the system in two directions: more instantiation functions and more special grammar constructs will widen the range of formats that can be described by specification files; and possibilities will be added for using C₂M as middleware system. We view the project as a first step toward a more empirical approach to language design, which will involve (among other things) user surveys.

Acknowledgements

The authors wish to acknowledge the many helpful comments from Judith Plümer (University of Osnabrück) and Andrei Malchanau (University of Twente).

Appendix A: Simple-chem ontology

```
<C2M-SPECIFICATION name="simple-chem"  
    type="ontology">
```

```
<TEXT>This example lacks documentation because the choices have been explained in the main text, section 3.3.</TEXT>
```

```
<ONTOLOGY>
```

```
molecule = atom-list, bond-list
```

```
atom-list = repeated( atom )
```

```
bond-list = repeated( bond )
```

```
atom = chemical-symbol, id, dr-coord-x, dr-coord-y, dr-coord-z
```

```
bond = id1, id2, bond-type
```

```
</ONTOLOGY>
```

```
</C2M-SPECIFICATION>
```

Appendix B: CT file format specification

```
<C2M-SPECIFICATION name="ct"  
    filetype="plaintext"  
    type="file-format">
```

```
<TEXT>Here, too, documentation has not been inserted because the main text provides sufficient background information to understand the choices made. See sections 3.4—3.6.</TEXT>
```

```
<READGRAM startsymbol="ct-file">
```

```
ct-file -> ct-first-line ct-numbers-line ct-atom-lines ct-bond-lines
ct-first-line -> line end-of-line
ct-numbers-line -> spaces nr-atom-lines spaces nr-bond-lines end-of-line
ct-atom-lines -> ct-atom-line+
ct-bond-lines -> ct-bond-line+
ct-atom-line -> spaces ct-coord-x spaces ct-coord-y spaces ct-coord-z spaces
    ct-chem-symbol end-of-line
ct-bond-line -> spaces ct-id1 spaces ct-id2 spaces ct-bond-type spaces
    ct-bond-order end-of-line
ct-coord-x -> float
ct-coord-y -> float
ct-coord-z -> float
ct-chem-symbol -> upper-case-letter
ct-chem-symbol -> upper-case-letter lower-case-letter
ct-id1 -> integer
ct-id2 -> integer
ct-bond-type -> integer
ct-bond-order -> integer
</READGRAM>
<SBREAD ontology="simple-chem"
    top-concept="molecule">
molecule =^ ct-file
atom-list =^ ct-atom-lines
bond-list =^ ct-bond-lines
dr-coord-x <- ct-coord-x
dr-coord-y <- ct-coord-y
dr-coord-z <- ct-coord-z
chemical-symbol <- ct-chem-symbol
id <- position(ct-atom-line, ct-atom-lines)
id1 <- ct-id1
id2 <- ct-id2
bond-type <- ct-bond-type
</SBREAD>
<WRITEGRAM start-symbol="ct-file">
ct-file -> "File generated by C2M" end-of-line
    space space nr-atom-lines space space nr-bond-lines end-of-line
```

```
fl-text(ct-atom-lines, ct-atom-line, nr-atom-lines)
fl-text(ct-bond-lines, ct-bond-line, nr-bond-lines)
ct-atom-line -> space space ct-coord-x space space ct-coord-y
                space space ct-coord-z space space ct-chem-symbol
                end-of-line
ct-bond-line -> space space ct-id1 space space ct-id2
                space space ct-bond-type space space ct-bond-type
                end-of-line
<TEXT>We just fill in the bond type where ChemDraw expects bond
order: this will not harm</TEXT>
</WRITEGRAM>
<SBWRITE>
<TEXT>This block is empty because as a matter of fact all semantic bindings in
the SBREAD-block are bijective.</TEXT>
</SBWRITE>
</C2M-SPECIFICATION>
```

Contact

Paul van der Vet and Eelco Mossel
Department of Computer Science, University of Twente
P.O. Box 217, 7500 AE Enschede, the Netherlands
Phone +31 53 489 3694, fax +31 53 489 3503
Email {vet,mossel}@cs.utwente.nl

Analyse der Qualität der multimedialen Lernumgebung „BioPrinz“

Josef Wiemeyer

Institut für Sportwissenschaft, TU Darmstadt
E-Mail wiemeyer@ifs.sport.tu-darmstadt.de

Abstract

The multimedia learning environment “BioPrinz“ has been developed and applied at the Institute of Sport Science in Darmstadt. This system serves as a complement for the basic course “Introduction to the biomechanics of sport” and the goal is to introduce the principles of biomechanics according to Hochmuth.

Currently, the second version of the learning environment consists of two parts: The first part contains information concerning the principles of biomechanics and the second part contains general additional information. The five main chapters of part 1 are structured into introduction, examples, animations and simulations, definitions and explanations, tasks, exercises and quiz.

The learning environment has been tested twice with comparatively small samples of students. Different methods of data acquisition have been applied: thinking aloud, log book, qualitative interviews, two ISO norm questionnaires, a multimedia questionnaire and a knowledge test (paper-and-pencil).

Overall, we found a good acceptance of the system. The analysis of the qualitative interviews and the ISO norm questionnaire showed several aspects for improvements. By means of the multimedia questionnaire we could find certain habits of the learners. The knowledge tests showed a trend for improvement of both performance and subjective confidence.

Einleitung

In einem interdisziplinären Lehrprojekt, das von der AG „Aktivierende Lernformen“ der TU Darmstadt finanziell unterstützt wurde, entwickelten Diplomstudierende der Sportwissenschaft im Sommersemester 2001 die Lernumgebung „BioPrinz“. Die Multimedia-Bausteine wurden im Rahmen der Lehrveranstaltung „Grafik/ Animation“ konzipiert und dann zur Lernumgebung integriert.

Die erste Anwendung erfolgte im Wintersemester 2001/2002 im Rahmen des Proseminars „Einführung in die Biomechanik“. Auf der Grundlage der ersten

Evaluation erfolgte eine vollständige Überarbeitung (Version 2.0) und Neugliederung. Die zweite Version wurde im Sommersemester 2002 eingesetzt und erneut evaluiert.

Als Entwicklungstools wurden primär die folgenden Programme eingesetzt: Macromedia Authorware, Macromedia Director, Microsoft Visual Basic, Java und Adobe Premiere.

In diesem Beitrag sollen kurz Aufbau und Inhalte der Lernumgebung dargestellt werden, um dann Vorgehen und Ergebnisse der (formativen) Evaluation des praktischen Einsatzes darzustellen.

Aufbau und Inhalte der Lernumgebung

Die Lernumgebung (Version 2.0) besteht aus zwei Teilen (s. Abbildung 1).



Abbildung 1: Screenshot der Benutzungsoberfläche von BioPrinz V.2.0

Zunächst werden im Teil 1 die biomechanischen Prinzipien (Anfangskraft, optimaler Beschleunigungsweg, Koordination der Teilimpulse, Gegenwirkung, Impulserhaltung) eingeführt, gefolgt von Glossar, Literatur und Links im Teil 2.

Jedes Kapitel im ersten Teil ist untergliedert in sieben Abschnitte:

- Ein Überblick gibt kompakte Informationen zum jeweiligen Prinzip.

- Beispiele zeigen die konkrete Anwendung des jeweiligen Prinzips auf Sportbewegungen.
- Animationen bzw. Simulationen veranschaulichen wichtige Aspekte, die mit dem jeweiligen Prinzip zusammenhängen. So können sich die Studierenden z.B. für eine von ihnen gewählte Kombination von Beschleunigungsweg und gewünschter Endgeschwindigkeit die erforderliche mittlere Beschleunigung und die maximale Beschleunigungsleistung anzeigen lassen. Sie können mit den Simulationsprogrammen selbständig oder nach Anweisung experimentieren (s. Abbildung 2).

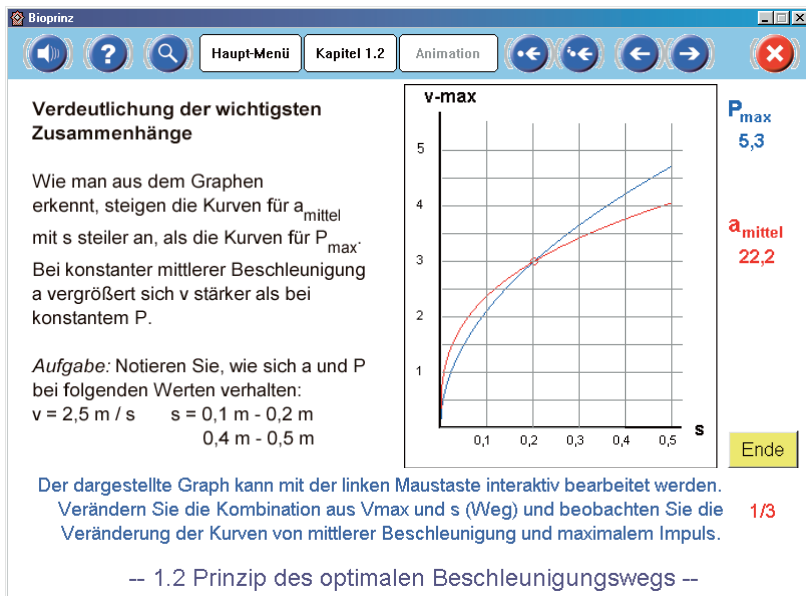



Abbildung 2: Simulationsaufgabe zum optimalen Beschleunigungsweg

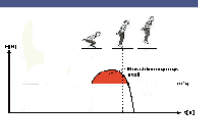
- Definitionen und Erklärungen liefern die eigentlichen Lerninformationen. Hier werden die einzelnen biomechanischen Prinzipien eingehend erläutert (in der Regel mit entsprechenden Grafiken).
- Aufgaben dienen der Bearbeitung ausgewählter Problemstellungen. Den Studierenden wird z.B. im Kapitel „Anfangskraft“ die Aufgabe gestellt, in der Turnhalle selbst Sprünge mit und ohne Ausholbewegung durchzuführen und die Ergebnisse zu protokollieren und auszuwerten.
- Übungen dienen der Vertiefung bzw. Festigung des Lernstoffes.

- Ein Quiz dient jeweils der Überprüfung des Wissens. Hier werden verschiedene Fragen- bzw. Aufgabentypen eingesetzt: Single-Choice-Fragen, Multiple-Choice-Fragen, Hot-Object-Fragen, Text-Fragen und Drag-and-drop-Fragen (s. Abbildung 3).

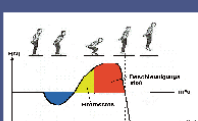
Kraftkurven zuordnen  **Frage 2 von 10**

Quiz zum Thema "Prinzip der Anfangskraft"

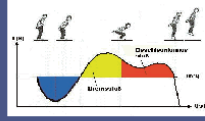
Ziehen Sie die Kraft-Kurven an die korrekten Positionen!



optimaler Sprung



zu starkes Bremsen



zu kleines Kappa

[Antwort prüfen](#)

Quiz zum Thema "Prinzip der Anfangskraft"

Abbildung 3: Screenshot einer Drag-and-Drop-Aufgabe von BioPrinz V2.0

Damit wird sicherlich nur ein Teil der vielfältigen, beim multimedialen Lernen einsetzbaren didaktischen Konzepte in dieser Lernumgebung umgesetzt (vgl. Kearsley, 2001).

Es bestehen verschiedene Möglichkeiten, sich den Lernstoff selbständig oder entsprechend der Anordnung der Kapitel zu erarbeiten. Die Benutzer können sich frei im System bewegen oder einfach sequentiell navigieren, indem sie von einer Seite zur nächsten gehen. Sie werden bei der Navigation durch verschiedene Funktionen unterstützt:

- Navigation von Seite zu Seite
- Navigation zur zuletzt besuchten Seite
- Navigation zu den einzelnen Kapiteln und Unterkapiteln (Aktivierung der Übersicht über die rechte Maustaste)
- Volltextsuche nach allen in den Lehrtexten verwendeten Begriffen

Evaluation – Fragestellungen und Vorgehen

Die Lernumgebung wurde mittlerweile zweimal eingesetzt und evaluiert.

Dabei wurden die folgenden Stichproben untersucht:

- Wintersemester 2000/2001: Die Lernumgebung „BioPrinz“ V. 1.0 wurde von 14 Sportstudierenden eingesetzt (Alter: $M = 23.8$ Jahre; Computererfahrung: $M = 6.25$ Jahre).
- Sommersemester 2002: Die Lernumgebung „BioPrinz“ V. 2.0 wurde von 8 Sportstudierenden eingesetzt.

Die folgenden Aspekte sollten erforscht werden:

- *Inhalte der Lernumgebung*: Insbesondere die Verständlichkeit und die Struktur der Lernumgebung wurden untersucht.
- *Formale Gestaltung der Lernumgebung*: Hier interessierte vor allem die Qualität der Dialoggestaltung nach ISO-Norm 9241-10 (Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Fehlertoleranz, Erwartungskonformität, Steuerbarkeit, Individualisierbarkeit, Lernförderlichkeit), der Einsatz von Farben, Schrift etc. sowie das Screendesign.
- *Nutzung der Lernumgebung*: Im Hinblick auf die Nutzung der Lernumgebung wurden u.a. Nutzungsverhalten (Wochentage, Tageszeiten, Dauer der Sitzungen), Nutzung der verschiedenen Navigationshilfen und Systemkomponenten, Einstellungen, Gebrauchstauglichkeit und Lernerfolg untersucht.

Es ging also bei den (formativen) Evaluationen nicht darum, Effektivitäts- oder Effizienzvergleiche anzustellen, sondern erstens Daten zur Verbesserung und zweitens Daten zur tatsächlichen Nutzung der Lernumgebung zu gewinnen. Insofern sind das Fehlen einer Kontrollgruppe und die unzureichende Erfassung potentieller Störvariablen, die oft als Qualitätsdefizit summativer Evaluationsstudien zum Multimedia-Lernen kritisiert werden (z.B. Kirkpatrick & Cuban, 1998; IHEP, 1999; Blömeke, 2003), hier nicht so gravierend.

Als Methoden wurden lautes Denken, Fragebögen (ISO-Norm-Fragebögen von Prümper & Anft, 1993 bzw. Willumeit, Hamborg & Gediga, 1997 und ein selbstentwickelter Multimedia-Fragebogen), Tagebuch, qualitative Interviews und Wissenstests eingesetzt (zu grundsätzlichen Möglichkeiten der Evaluation vgl. Wiemeyer, 2003).

Evaluation - Ergebnisse

Als wesentliche Ergebnisse erbrachten die Evaluationen die folgenden Erkenntnisse.

Die Inhalte der Lernumgebung waren - bis auf wenige Ausnahmen - gut verständlich (s. Abbildung 4). Nur bei einer Simulation ergaben sich Verständnisprobleme. Auch allgemeine Akzeptanz, Form, Nutzung und Lernen mit der Lernumgebung wurden im Durchschnitt als gut bewertet (s. Abbildung 4).

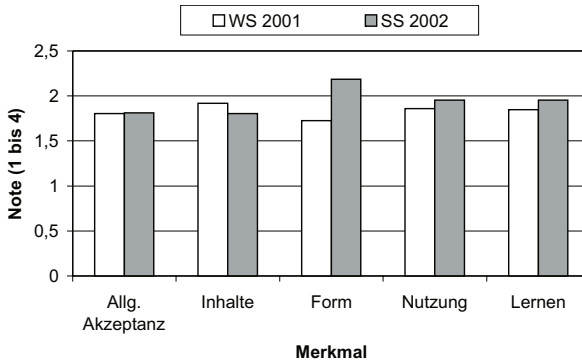


Abbildung 4: Allgemeine Bewertung von BioPrinz V.2.0 (WS 2001: N=14; SS 2002: N=8)

Die Gliederung der Lernumgebung war klar und übersichtlich, die Navigationsmöglichkeiten wurden als gut bewertet. Besonders positiv wurden die Integration von Aufgaben in den Text und der Einsatz von unterschiedlichen Farben bewertet (s. Abbildung 5).

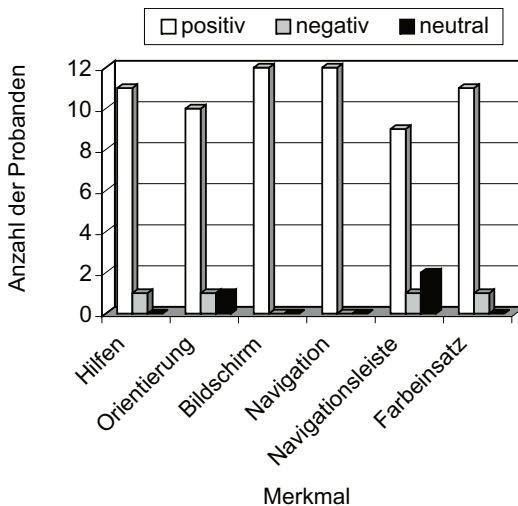


Abbildung 5: Ergebnisse der qualitativen Interviews im Wintersemester 2001/02

Die Lernumgebung wurde ein- bis zweimal in der Woche benutzt. Die durchschnittliche Nutzungsdauer betrug 32 bis 36 Minuten pro Sitzung. Die bevorzugte Nutzungszeit war 12 bis 16 Uhr im Wintersemester 2001 und 16 bis 20 Uhr im Sommersemester 2002.

Die Studierenden bevorzugten eine sequentielle Navigation (Seite vor/ zurück). Am zweithäufigsten wurden Rücksprungbutton und Menübuttons genutzt (s. Abbildung 6).

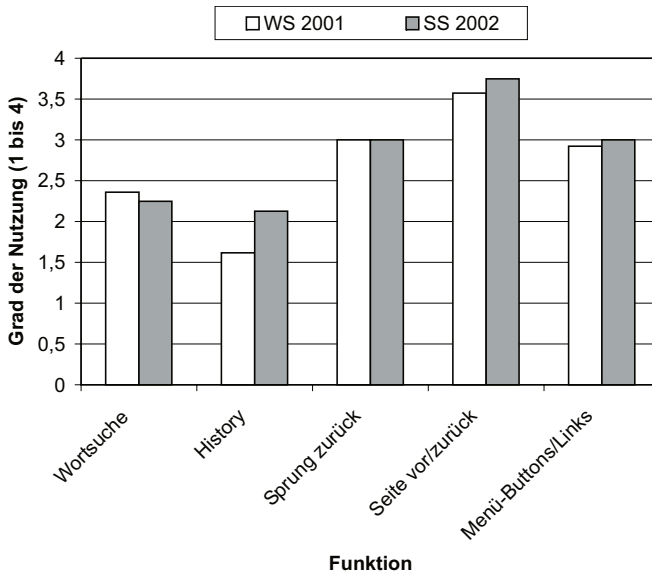


Abbildung 6: Nutzung der Navigationsfunktionen von BioPrinz

Hinsichtlich der Dialoggestaltung nach ISO Norm 9241-10 ergab sich ein Zustimmungsgrad zwischen 68 und 86% - mit einer Ausnahme (s. Abbildung 7). Die deutlich abweichenden Ergebnisse zum ISO-Item „Individualisierbarkeit“ spiegeln die Tatsache wider, dass die Lernumgebung keine individuelle Anpassung ermöglicht.

Weiterhin wird aus Abbildung 6 deutlich, dass sich die Bewertung von fünf der sieben Items mehr oder weniger deutlich von Version 1 zu Version 2 verbessert.

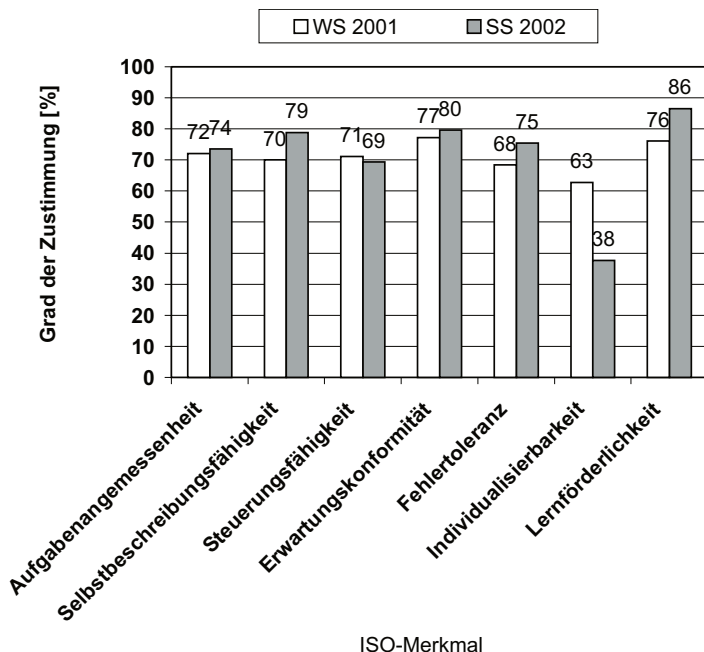


Abbildung 7: Ergebnisse zur Dialoggestaltung nach ISO Norm 9241-10

Die vor und nach dem Einsatz der Version 2.0 durchgeführten Wissenstests ergaben eine deutliche Zunahme der Quote der richtig gelösten Aufgaben und der subjektiven Sicherheit bei der Aufgabenlösung (siehe Abbildung 8). Allerdings können diese Ergebnisse lediglich als Trend interpretiert werden, da erstens die Zahl der Studierenden sehr gering ist und zweitens eine Wechselwirkung zwischen dem Einfluss der Lernumgebung und dem Einfluss des Unterrichts im Seminar besteht. Vorliegende Befunde zeigen, dass gerade die Integration von Multimedia in den traditionellen Unterricht besonders effektiv ist (vgl. Roblyer, Castine & King, 1987).

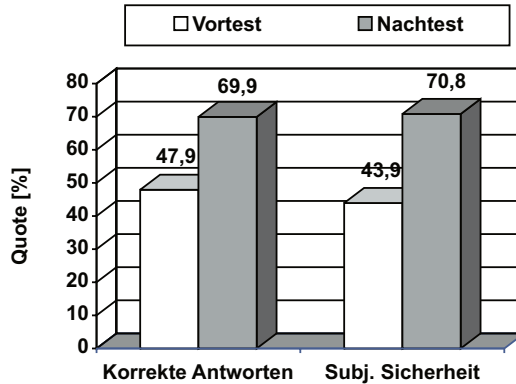


Abbildung 8: Ergebnisse der Wissenstests vor und nach dem Einsatz von „Bio-Prinz“ im Sommersemester 2002

Zusammenfassung und Ausblick

Insgesamt zeigen die Evaluationen, dass das multimediale Lernen mit der multimedialen Lernumgebung „BioPrinz“ von den Studierenden als eine Bereicherung des universitären Lernens eingeschätzt wurde. Insbesondere die qualitativen formativen Evaluationsmethoden (lautes Denken, qualitatives Interview) lieferten aussagekräftige Ergebnisse, die zur Verbesserung der Lernumgebung eingesetzt werden können.

Die Lernumgebung wird zum Sommersemester 2003 nochmals gründlich überarbeitet und dann zum dritten Mal im Proseminar „Einführung in die Biomechanik“ eingesetzt und evaluiert.

Literatur

- Blömeke, S. (2003). Lehren und Lernen mit neuen Medien - Forschungsstand und Forschungsperspektiven. *Unterrichtswissenschaft*, 31 (1), 57-82.
- Institute for Higher Education Policy [IHEP] (1999). What's the difference? Abgerufen am 8.5.2003 von <http://www.nea.org/he/abouthe/diseddif.pdf>
- Kearsley, G. (2001). The theories. Abgerufen am 5.9.2002 von <http://tip.psychology.org/theories.html>
- Kirkpatrick, H. & Cuban, L. (1998). Computers make kids smarter - right? *Technos Quarterly*, 7 (2), 1-11.
- Prümper, J. & Anft, M. (1993). Fragebogen ISO-NORM 9241/10. Abgerufen 17.6.2002 von <http://141.90.2.11/ergo-online/Software/ISO-Fragebogen.htm>

- Roblyer, M.D., Castine, W.H. & King, F.J. (1988). Assessing the impact of computer-based instruction. New York: Haworth Press.
- Wiemeyer, J. (2003). Evaluation of multimedia programs in sport science education. *International Journal of Computer Science in Sport*, (Special edition 1), 41-50.
- Willumeit, H., Hamborg, K.C. & Gediga, G. (1997). IsometricsL. Fragebogen zur Evaluation von graphischen Benutzungsschnittstellen. Osnabrück: Fachbereich Psychologie.

Ziele des erstmals von der Initiative Information und Kommunikation wissenschaftlicher Fachgesellschaften (IuK) zu ihrer Frühjahrstagung herausgegebenen referierten Tagungsbandes sind intensiver Diskurs über internetgestützte Techniken in der Kommunikation wissenschaftlicher Gruppen und Gemeinschaften sowie Momentaufnahme der Entwicklung digitaler Bibliotheken in der Bundesrepublik.

Der Tagungsband vertieft die Webproceedings, die Dokumentation aller Vorträge enthalten und unter <http://www.iwi-iuk.org/iuk2003/> zugänglich sind.



InformationsZentrum
Sozialwissenschaften

der Arbeitsgemeinschaft
Sozialwissenschaftlicher Institute e.V.

Lennéstraße 30 • D-53113 Bonn
Telefon 02 28 / 22 81 - 0
Telefax 02 28 / 22 81 - 120

GESIS

Das IZ ist Mitglied der
Gesellschaft Sozialwissenschaftlicher
Infrastruktureinrichtungen e.V.

Die GESIS ist Mitglied der
Leibniz-Gemeinschaft.

ISBN 3-8206-0145-7
EUR 25,-